Naval Research Laboratory

Monterey, CA 93943-5502

NRL/MR/7531/97/7232

Three Dimensional Covariance Functions: Real Data

RICHARD FRANKE

Atmospheric Dynamics & Prediction Branch
Marine Meteorology Division

October 1997

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. Agency Use Only *(Leave Blank)*.	2. Report Date.	3. Report Type and Dates Covered.
	October 1997	Final

4. Title and Subtitle.
Three Dimensional Covariance Functions: Real Data

5. Funding Numbers.
PE 0602435N
PN BE 35-2-19

6. Author(s).
Richard Franke

7. Performing Organization Name(s) and Address(es).
Naval Research Laboratory
Marine Meteorology Division
Monterey, CA 93943-5502

8. Performing Organization Reporting Number.
NRL/MR/7531 - 97- - 7232

8. Sponsoring/Monitoring Agency Name(s) and Address(es).
Naval Research Laboratory (NRL)

10. Sponsoring/Monitoring Agency Report Number.

11. Supplementary Notes.

12a. Distribution /Availability Statement.
Approved for public release; distribution unlimited

12b. Distribution Code.

13. Abstract *(Maximum 200 words)*.

Height innovation data for a two-month period from NOGAPS was analyzed to obtain height prediction and observation error covariances. Different methods of weighting the data in least squares approximations were investigated using the second order autoregressive correlation function, both with and without an additive constant (varying with pressure level). Based on the properties of the derived covariance matrices and its parameters, the SOAR without an additive constant was used for the horizontal approximations. The vertical correlations were fit using a combination of SOAR plus and additive constant and a transformation of the logP coordinate to another coordinate to achieve a best fit. The resulting three dimensional approximation is partially separable, being the product of the horizontal covariance function (which depends on height) and the vertical correlation function. Figures demonstration various aspects of the process and the results are given.

14. Subject Terms.
Three Dimensional Covariance Functions: Real Data

15. Number of Pages. 46

16. Price Code.

17. Security Classification of Report.	18. Security Classification of This Page.	19. Security Classification of Abstract.	20. Limitation of abstract.
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	Same as report

Contents

1. Introduction

This report details the process and results of investigating approximations to the covariance structure of the predicted and observed pressure height errors obtained from a two month history of the Navy Operational Global Atmospheric Prediction System (NOGAPS) that is used by the Fleet Numerical Meteorological and Oceanographic Center (FNMOC) for daily atmospheric prediction. The data covered the period October-November 1996, and consisted of innovation data (predicted height of pressure levels minus radiosonde measurements) at times 00Z and 12Z. The radiosonde data had been subjected to and passed the routine quality control measures applied by the system. For the purposes of this study, all valid data covering a portion of North America was included, that between 25° to 65° North and 70° to 130° West. The data set was comprised of height "errors" at (up to) 16 pressure levels over the 61 day period for 89 stations at 00Z, and at 92 stations at 12Z. To be specific, the 16 pressure levels are 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, and 10 mb. The two times were always treated individually. The mean value of the data at each level for each station and each time was removed from the raw data to obtain the data to be processed, and it is this resulting data that is used in all instances.

This investigation is very likely similar to others that have been reported in the literature since the advent of numerical weather prediction. A partial list includes Julian and Thiébaux [8], Lönnberg and Hollingsworth [9], Hollingsworth and Lönnberg [7], Barker [1], Rabier and McNally [13], Daley [3], Bartello and Mitchell [2], Da Silva, et al. [4], and Mitchell, et al. [10] and [11]. The data is from a different source, of course, but in addition, this study incorporated extensive computations to investigate various aspects and parameters of the spatial covariance fitting procedure, along with an attempt to justify the appropriateness of the final form of the approximation. This investigation follows one reported in a prior technical report by the author [5].

The goal of the investigation was to obtain a consistent and valid approximation for the covariances of the height-height errors. Certain assumptions must be made about the data. In addition to the usual assumptions with respect to time, it was assumed to be homogeneous for any given pressure level, but having properties that vary with height. The final approximation

is assumed to be partially separable in that the correlation function is a product of a vertical correlation function and a horizontal correlation function. The horizontal correlation function depends on height, but the vertical correlation function is taken to be independent of horizontal position. The basic tool is the approximation of the covariance data as a function of distance for each pair of levels. This yields an intercept (the "prediction" error covariance), which when subtracted from the zero distance empirical covariance data, yields the "observation" error covariance.

The data consists of time histories of data at a given level for 61 days at some (89 or 92, depending on the time of day) stations. For the moment let the data be represented by $e_{i,l,t}$, where i,l,t represent station, pressure level, and time. The empirical covariance data $\overline{e_{i,l,t}e_{j,l,t}} = \frac{1}{N_{i,j,l}}\sum_t e_{i,l,t}e_{j,l,t}$ is formed. Here the overbar indicates the average value, $N_{i,j,l}$ is the number of times at which both errors exist for the same time at both stations of the pair at the given level. The value of $E_{i,j,l} = \overline{e_{i,l,t}e_{j,l,t}}$ is tabulated, along with the distance between stations. For this study, the distance was taken to be in "pseudoradians", that is $d = \sqrt{(\Delta lat)^2+(\cos 40^o \Delta long)^2}$. This corresponds to distance on a latitude-longitude plane at 40° latitude. This does not preserve real distances, especially far away from 40°, but it does avoid complications with covariance functions on the sphere rather than in the plane. More will be said on this later. The time averaged data then consists of points $(d_{i,j}, E_{i,j,l})$, along with $N_{i,j,l}$ for all possible pairs of stations. All of the previous processing was performed using Fortran programs. Subsequent processing and plotting was all performed using Matlab[1]. The data is then collected into bins of width 0.01 radians, with boundaries at [0,0.0005, 0.0105,0 .0205, … , 0.4105], a total of 42 bins, although there is never any data in the second bin (distances in the [0.0005,0.0105) range). The abscissa and ordinate values of the station pairs in each bin are averaged to obtain the data $\{(\hat{d}_n, \hat{E}_n)\}_{n=0}^{41}$ that is to be fit. Note that for simplicity the dependence on level has been suppressed in this notation. In addition, N_n, the

[1] ® The MathWorks, Natick, MA

number of total terms in the time average for each bin has been saved for possible subsequent use.

The following sections will detail the process that was carried out, including all the variations and reasons for decisions as to which options seem to yield the best results. Plots are included throughout for illustrative purposes. The sections are basically in "chronological" order. Section 2 discusses the (horizontal) spatial covariance fitting, the choice of approximation and different weights used, as well as the details of obtaining the empirical covariance data for the prediction and observation errors. In what follows, errors will be interpreted as height errors unless termed otherwise. Section 3 deals with the vertical correlation function and obtaining an approximate covariance matrix for the prediction and observation errors. In Section 4 the estimation of spatial parameters in the covariance function approximation is treated, and the form of the height prediction error covariance function and how well it approximates the raw data is shown. Section 5 discusses the positive definteness aspects of the approximation. The final section gives a summary of the work and points out some possible future work.

2. Spatial Covariance Approximation

The choice of function that was made for fitting the data $\{(\hat{d}_n, \hat{E}_n)\}$ obtained above is a special case of the second order autoregressive (SOAR) correlation function (plus a constant) $F(d;a,b,c) = (1-c)\left(\cos ad + \frac{b\sin ad}{a}\right)e^{-bd} + c$, which in the limit as $a \to 0$ becomes

$$F(d;b,c) = (1-|c|)(1+bd)e^{-bd} + |c| \qquad (1)$$

when the absolute value of c is taken to emphasize that it must be non-negative. The absolute value will be dropped in further discussion with the assumption $c \geq 0$. In some cases the value $c = 0$ is imposed, yielding

$$F(d;b) = (1+bd)e^{-bd}\ . \qquad (2)$$

This function is widely known to be positive definite in the plane for all values of the parameters and is in current use in the present NOGAPS (see Goerss and Phoebus,[6]). Using the standard minimization package fmins in Matlab, the parameters C, b, and c are found that minimize $\sum_{n>0} W_n\left(CF(\hat{d}_n;b,c) - \hat{E}_n\right)^2$. This yields the weighted (weights W_n) least squares approximation to the binned data, and the corresponding values of C, b, and c. The choice of weights will be discussed later. The value of C is the intercept of the curve approximating the spatial covariance of the prediction errors at the given level and represents the variance of the prediction errors, while b is a scaling parameter, and c is an additive parameter. Because the observation errors for different radiosondes are assumed independent, their covariance is zero. However, the variance of the observation errors for a particular level and station does enter into the zero distance data, so the variance of the observation error is estimated by $C_O = \hat{E}_0 - C$. This process gives an estimate of the variances of the prediction and observation errors at the various levels.

What remains to be estimated are the inter-level values for the covariances of prediction and observation errors. This can be done in at least two ways, and calculations were carried out to investigate both approaches. The straightforward way is to form the inter-level empirical covariances, $\overline{e_{i,l,t}e_{j,k,t}} = \frac{1}{N_{i,j,l,k}}\sum_t e_{i,l,t}e_{j,k,t}$, and in analogy with what was described above, obtain the data $(d_{i,j}, E_{i,j,l,k})$, and subsequently the binned data $\{(\hat{d}_n, \hat{E}_n)\}$ for each pair of levels $l \neq k$, which is then fit in the same way. The same SOAR covariance approximation was used, although as we note later, a different approximation probably should be used. The alternate way of calculating the covariances between levels is to calculate the empirical thickness (differences in the values between two levels at the same station) covariance data and process it in the same way. This then yields the estimated values of the variance of the thickness prediction errors and the thickness observation errors for the l to k level thickness by the same difference process as for the height errors.

There is a simple relationship between the thickness error variances and the height error covariances. Letting $D_{l,k}$ represent the variance of the thickness errors (at zero distance) and $C_{l,k}$ the covariance of the errors at levels l and k, with v_l and v_k the data for levels l and k, we have $D_{l,k} = \overline{(v_l - v_k)(v_l - v_k)} = \overline{v_l v_l} - 2\overline{v_l v_k} + \overline{v_k v_k} = C_{l,l} - 2C_{k,l} + C_{k,k}$. Rearrangement then gives

$$C_{l,k} = \tfrac{1}{2}(C_{l,l} + C_{k,k} - D_{l,k}) . \qquad (3)$$

The same formula holds for the observation error covariances, simply substituting observation values $C_{O,l,k}$, $C_{O,l,l}$, $C_{O,k,k}$, and $D_{O,l,k}$ in place of the corresponding prediction values.

The results of using various weights to calculate the spatial covariance approximations will be discussed now. First we note that the height-height calculations resulted in data that is not well approximated by the assumed SOAR function. Because of this, the inter-level fits, independent of the weight, tended to be ill-behaved for several height pairs, giving results that were not reasonable (although corresponding to what was probably a good minimum value of the objective function). Therefore we focus our discussion here only on the various weights and parameter options used in fitting the spatial covariance for the prediction errors for each level, and for all inter-level thickness prediction errors. Some properties of the prediction error and observation error covariances are given in tabular form.

The table reveals that all cases have at least one negative eigenvalue. It is noted that the deletion of levels 15 and 16 from the matrices generally eliminated negative eigenvalues, but not always. The use of the additive constant, c, sometimes yields significantly better fits to the data. Although the additive constant generally tended to be zero for lower and upper level single height-height fits, it took on values as big as 0.15 at 12Z, and as big as 0.30 at 00Z at intermediate levels. A little numerical experimentation confirms that the three dimensional "covariance" function will almost certainly be indefinite if the additive constant is allowed to vary. As an alternative to the additive constant, Mitchell, et al. [10] added a second SOAR term using a fixed linear combination of the two SOAR terms, with a fixed relation between

Nr	Weight	Function	minimum C eigenvalue		minimum C_0 eigenvalue	
			00Z	12Z	00Z	12Z
1	N_n^2	(1)	-13.61	-5.39	-10.20	0.32
2	N_n	(1)	-8.66	-0.08	-6.20	2.79
3	$N_n / \bar{d}_n^2$	(1)	-63.70	-2.16	-102	-6.51
4	1	(1)	-27.78	-19.21	-56.65	-1.78
5	$N_n / \bar{d}_n$	(1)	-10.07	0.57	-9.27	2.60
6	N_n	(2)	-11.57	0.06	-6.39	3.85
7	1	(2)	-18.10	-16.73	-35.92	-1.30
8	$N_n / \bar{d}_n$	(2)	-10.90	0.19	-7.76	3.59
9	$N_n, n \le 20$ $0, n > 20$	(2)	-11.59	-0.40	-5.65	-0.10

Table 1: Eigenvalue properties of the covariance matrices.

the scaling parameters. It is not known how this affects positive definiteness when the parameter varies by level. Because of the wide variation of the additive value with level and its negative connotation in probability theory, it was decided to only consider (2) as the spatial covariance function to be used further in this study. Based on this and some admittedly subjective criteria, including the behavior of the height-height correlation distance and the transformation used to approximate it (see Section 4), it was decided that the least objectionable behavior corresponds using the weight approximation from line 9, although that in line 6 is also good. It is noteworthy that using the weight $W_n = 1$ (line 7, and 4) gives rather poor performance in terms of the eigenvalues of the matrices, and in the vertical approximations discussed in Section 4.

In order to see samples of the behavior exhibited by the data, and later by the approximations that are made, some examples are given here, and will be continued through the report. The data cloud (covariances between all station pairs) before binning is shown for time 12Z at the 500 mb pressure level in Figure 1. In Figure 2 the data is shown after binning,

time 12Z at the 500 mb pressure level in Figure 1. In Figure 2 the data is shown after binning, along with the approximating function of the form (2). Again at time 12Z, and for 500-400 mb, 500-200 mb, and 850-300 mb thicknesses, the binned data and the approximating functions are shown in Figures 3-5.

Having computed all the single height-height error parameters and all possible inter-level thickness-thickness error parameters, we have the following 16x16 matrices. $D_{l,k}$ - for $l \neq k$, the variances of the l to k thickness prediction errors. $C_{l,k}$ - for $l = k$ the variances of the height prediction errors, and for $l \neq k$, computed using equation (3). $D_{O,l,k}$ - for $l \neq k$, the variances of the l to k thickness observation errors. $C_{O,l,k}$ - for $l = k$ the variances of the height observation errors, and for $l \neq k$, computed using equation (3). $b_{l,k}$ - for $l = k$ the distance scaling parameters for height prediction error, and for $l \neq k$ the distance scaling parameters for the thickness prediction errors. In some sense it is more convenient to think in terms of a "correlation distance" parameter by taking the reciprocal of the $b_{l,k}$, so we introduce $\beta_{l,k} = 1/b_{l,k}$. In the case of fitting with equation (1), there would also be the additive constant matrix, $c_{l,k}$ - for $l = k$ the additive constant parameters for height prediction error, and for $l \neq k$ the additive constant parameters for the thickness prediction errors.

Given this data, it is now possible to infer the approximation to the inter-level height prediction errors. By carrying along the possibility of nonzero distances in the process leading to equation (3), we find that

$$C_{l,k}(d) = \tfrac{1}{2}\left(C_{l,l}(1+d/\beta_{l,l})e^{-d/\beta_{l,l}} + C_{k,k}(1+d/\beta_{k,k})e^{-d/\beta_{k,k}} - D_{l,k}(1+d/\beta_{l,k})e^{-d/\beta_{l,k}}\right). \quad (4)$$

This expression allows us to explore the prediction height-height approximations as implied by equation 4. These curves and the binned data are shown in Figures 6-8 for the same time and pressure levels as for Figures 3-5. It should be noted that the data may not be exactly the same since for the thickness covariance at nonzero distance four measurements must exist, while only two need exist for a height covariance. It is noted that the 850-300 mb covariance

function is very different from the SOAR function, but does not seem to be an unreasonable approximation to the data. In all cases it should be noted that while the data is shown out to a distance of about 0.40 radians, the fitting process did not use any points beyond 0.20 radians.

3. Vertical Covariance Approximation

Whichever of the two procedures are followed, the results are approximate covariance matrices C for the prediction errors, and C_O for the observation errors. In principle, these matrices are positive definite, however because of inadequate data (especially for small distances), and differing amounts of data, they may in fact be indefinite as we have noted in the previous section. The next step is to fit a vertical correlation function to the empirical correlation matrices $\tilde{C}$ and $\tilde{C}_O$ obtained from C and C_O by dividing each row and column by the square root of the diagonal element (that is, normalizing each row and column by the standard deviation). Call the diagonal matrices whose diagonals are the square roots of the diagonals of those matrices, σ and σ_o, respectively. The diagonal entries are the standard deviations of the prediction errors and the observation errors at the various pressure levels, respectively. It is now proposed to fit the correlation data in the manner initiated by Sampson and Guttorp [14] and described in more detail in a previous report [5]. Letting the P_l values represent pressure levels, it is desired to simultaneously find a transformation $h = g(\log P)$ with $h_l = g(\log P_l)$ and parameter γ so that $\sum_{l=1}^{16}\sum_{k=l}^{16}\left((1-|\gamma|)(1+(|h_l-h_k|)e^{-|h_l-h_k|}+|\gamma|-\tilde{C}_{l,k}\right)^2$ is minimized (as before we drop the absolute value, assuming $\gamma \geq 0$). Again, the standard routine fmins from Matlab was used. The approximate covariance matrix for the prediction errors then becomes $\sigma\left[(1-|\gamma|)(1+|h_l-h_k|)e^{-|h_l-h_k|}+|\gamma|\right)\right]\sigma$. This matrix is guaranteed to be positive definite and matches exactly the variances (diagonal elements) at each level. An interpolation must be done for intermediate levels, if necessary. Alternately, an approximation could be carried out to obtain a functional form for σ, but it has not been done at this time. The process is repeated for the observation error.

With the approximate covariance matrix for the prediction errors given above, the implied expression for the level l to k thickness errors then becomes $D'_{l,k} = \sigma_l^2 + \sigma_k^2 - 2\sigma_l\sigma_k\,[(1-|\gamma|)(1+\delta h)e^{-\delta h} + |\gamma|]$, by solving equation (3) for $D_{l,k}$, and calling the approximation $D'_{l,k}$.

Figure 9 shows the prediction error correlation data as a function of logP distance at time 12Z, and Figure 10 shows the same data, along with the approximating function in $g(\log P)$ distance. Figure 11 shows the $h = g(\log P)$ transformation. The resulting vertical correlation function and the raw data from $\tilde{C}$ are shown in Figure 12. For comparison, these calculated correlation curves are shown together with the presently used vertical correlation curves (see [6]) in Figure 13. Note that the presently used curves are considerably more narrow that those obtained here. The figures corresponding to Figures 9-12 for the height observation error are given in Figures 14-17 (the analog to Figure 13 is not given since observation errors are presently assumed to be vertically uncorrelated).

Recall that the standard deviation of the prediction error at the various levels is given by the diagonal of σ, and of the observation error by the diagonal of σ_o, with the standard deviation of the total error given by $\sigma_T = \sqrt{\sigma^2 + \sigma_o^2}$. The standard deviation curves for 12Z are given in Figure 18. The solid curves are those obtained here, while the dashed curves are those presently used in NOGAPS. The standard deviations are marked with "p", "o", and "t" for prediction, observation, and total error. It is noteworthy that the presently obtained curve for prediction error lies mostly to the left of the curve for observation error, while the opposite is true for the currently used values. The corresponding plot for 00Z is given in Figure 19. Note that the curves at the two times are very close to being the same, indicating the standard deviations of the prediction, observation, and total errors seem to be independent of the time of day, a reasonable result.

Because the weight $W_n = 1$ may be routinely used in obtaining covariance function approximations, some additional results will now be cited for that weighting. The figures corresponding to Figures 9, 10, and 11 for the weight $W_n = 1$ are given in Figures 20, 21, and 22. Note the relatively less scatter in Figures 9 and 10, compared with Figures 20 and 21, and

the relatively benign behavior of the transformation in 11 compared to 22. The results corresponding to Figures 18 and 19 are given in Figures 23 and 24. These differ considerably from each other at the higher levels, and also from both of 18-19. A symptom of these inconsistencies was noted earlier in terms of the eigenvalues of the implied covariance matrices. Some possible reasons for poor behavior of approximations computed using $W_n = 1$ can also be seen in Figure 25, a bar chart showing the number of terms from the time averaging that fall into each bin for the 500 mb level. This number is the term N_n used in the weighted least squares. The number of terms for bins at short distances is relatively small and such data should not be given so much credence as to have the same weight as data in better represented bins.

4. The Correlation Distance Parameter

The spatial scaling parameter $b_{l,k}$ truly seems to depend on both l and k, however it is desirable to simplify this dependence as well as to smooth it. The scaling parameter for the prediction height-height errors, broadly speaking, increases with height initially, then decreases, and finally increases once more. Of course, the computed values are not quite so well behaved. It seems be more intuitive to treat the reciprocal since this represents a "correlation distance", that is, it would be the distance at which the exponent takes on the value one. Therefore we convert from $b_{l,k}$ to $\beta_{l,k} = 1/b_{l,k}$. For $k = l$ the data at pressure levels from 1000 to 30 mb was approximated by a cubic equation in $\log P$, with the constraint that the curve pass through 0.08 at 10 mb. The latter was somewhat arbitrary in that the values obtained during the fitting process were considerably larger (12Z) and smaller (00Z) than this value at the 10 mb level . However, the relatively small amount of data at higher levels, along with the idea that the correlation distance ought to be smaller at higher levels prompted this device to assure "reasonable" behavior. The computed correlation distances and the approximating curves are shown in Figure 26. Given the consistency between the standard deviations at 00Z and 12Z, as noted earlier, it is somewhat puzzling that the

correlation distances turn out to be considerably different for the two times, indicating a different spatial structure for the prediction errors at the two times.

For $k \neq l$ there is considerable variation with level, and it is unclear how the correlation distance should be approximated. In the interests of keeping the approximation simple and not introducing too much variation into it (with the idea of enhancing the maintenance of positive definiteness of the approximation), it was decided to use the approximation $\beta_{l,k} = \sqrt{\beta_{l,l}\beta_{k,k}}$. With this approximation incorporated into the expression for $D'_{l,k}$, and substituting all approximations into equation (4), the final approximation to the prediction errors is given by

$$C_{l,k}(d,\delta h) = \frac{1}{2}\begin{pmatrix} \sigma_l^2(1+d/\beta_{l,l})e^{-d/\beta_{l,l}} + \sigma_k^2(1+d/\beta_{k,k})e^{-d/\beta_{k,k}} - \\ \left(\sigma_l^2 + \sigma_k^2 - \sigma_l\sigma_k\left[(1-\gamma)(1+\delta h)e^{-\delta h} + \gamma\right]\right)(1+d/\beta_{l,k})e^{-d/\beta_{k,l}} \end{pmatrix} \qquad (5)$$

Note, however, that despite the generality of notation on the left side, that this really is quite specific in that it represents the covariance between the prediction errors at levels l and k, where δh is the distance between levels in the transformed coordinate. If the covariance between arbitrary pressure levels is needed, the it is necessary to approximate (interpolate) for the standard deviations of the prediction error at the two levels, given by σ_l and σ_k in equation (5), along with the values of the β s.

At this point the final approximations corresponding to Figures 2-8 can be computed and plotted. These are shown in Figures 27-33, respectively. Some observations can be made. The 500 mb height error approximation is very good, as can be seen by comparing Figures 2 and 27. This follows because the variance is correct, and the correlation distance approximation is good. The 500 to 400 mb thickness error approximation in Figure 28 is not as good (compare with Figure 3), reflecting the poor approximation of the vertical correlation (this can be seen in Figure 12). On the other hand, the error approximation in Figure 31 (compare with Figure 5) is quite reasonable. All approximations of the 500 to 200 mb data seem reasonable, as seen by comparing Figures 4 and 29, and 7 and 32. For the 850 to 300 mb thickness, the approximation in Figure 30 is somewhat poorer than that shown in Figure 5,

with the intercept being depressed while the correlation distance is considerably larger in the approximation. The corresponding figures for the height errors show marked differences. The initial increase in the covariance with distance shown in Figure 8 is completely missing from Figure 33. This is probably a result of the correlation distance approximation being quite different from the computed one for the 850 to 300 mb thickness data. While examples of the persistence of initial increase of covariance after the correlation distance approximation do exist, for the most part the behavior is suppressed. In several instances, the covariance curve is flattened for small distances.

As a final comparison with the results obtained using $W_n = 1$, the plot of correlation distance versus pressure level for this case is given in Figure 34. While the cubic fit was not used since we didn't pursue this option, it is shown for comparison. The notable behavior is for 70, 50, and 30 mb levels at 00Z, for which the correlation distances are quite large. By contrast, the correlation distances for 12Z are rather well-behaved, although considerably smaller at the higher levels than obtained for the case chosen, on line 9 of Table 1. The two cubic fits vary considerably from each other and from the results shown in Figure 26.

The above approximations to the spatial correlation function are compared with the present NOGAPS spatial correlation function in Figure 35 for 12Z. The two solid curves give the SOAR function for the largest and smallest correlation distances using the cubic approximation as above, while the dashed line gives the correlation function (corresponding to a correlation distance of about 0.0604) from [6], which includes an additive constant of 0.1.

5. Positive Definiteness of the Approximation

While it is important for theoretical purposes to maintain the positive definiteness of the function approximating the covariance, it is imperative from the computational point of view. The use of Cholesky decomposition with its inherent stability require a positive definite matrix. While it is true that the addition of the observation error covariance into the matrix will tend to stabilize the computation (that is, it tends to decrease the condition number of the matrix), the importance of positive definiteness can scarcely be overstated.

For homogeneous processes, the key result is Bochner's theorem (see [12]). The essence of the theorem is that the Fourier transform of the (spatial) covariance function for the process is non-negative. There is a converse part of the theorem, as well. Unfortunately, as noted in the precursor report [5], the approximations pursued here are not homogeneous. Nonetheless, since the approximation is partially separable, some observations can be made. In particular, the Fourier transform in the horizontal can be evaluated explicitly,

obtaining $\dfrac{3\beta^2(h,h+\delta h)\sqrt{r}}{\sqrt{(1+r^2\beta^2(h,h+\delta h))^5}}(1+\delta h)e^{-\delta h}$. In the previous section, we took

$\beta(h,h+\delta h)=\sqrt{\beta(h)\beta(h+\delta h)}$, where the single argument β is the correlation distance for the pressure height error as a function of h while the two argument β is the correlation distance for thickness h to $h+\delta h$. The Fourier transform is easily seen to be positive for all values of the vertical coordinate h, the vertical distance parameter δh, and the horizontal (radial) coordinate in the Fourier transform space, r. Even for fixed values of h the Fourier transform in δh is not readily expressible in closed form. Some experiments were performed to calculate the Fourier transform numerically, looking particularly at small values of the transform parameter. No questionable results were obtained, but such experiments are inconclusive.

A secondary approach to the positive definiteness problem is to look at the assumed covariance function and attempt to find configurations of points (stations) for which the function does not lead to a positive definite matrix. To this end, the following experiment was conducted. A (pseudo) random number varying between 1 and 11 gave the number of stations. These were at random locations with latitudes varying by 0.25 radians and longitudes varying by 0.20 radians. The number of valid data levels varied (randomly) for each station between 8 and 16 (with data at levels 1 to that number). The "covariance" matrix was formed and the eigenvalues computed. In 6000 repetitions for each of the 00Z and 12Z cases, no negative eigenvalues were found. The size of the system could vary from 8 (unlikely) to 176 (even more unlikely; maximum size observed was around 150). For another 6000 repetitions using the 12Z data, the location of the stations were allowed to vary by 0.5 radians in both latitude and longitude, with no negative eigenvalues being found. Because

adverse behavior of the eigenvalues is likely to occur because of points within the correlation distance (the function becomes convex for greater distances) of each other, increasing the size of the region in which the stations lie is unlikely to lead to an indefinite matrix unless the smaller region did as well.

While the above procedures do not prove positive definiteness of the approximating function(s), it seems very unlikely they are indefinite. It should be further noted that the pressure levels at which we have tested are limited to the 1000 to 10 mb range, and it is probably true that the correlation distance function (as stated) would lead to indefinite matrices unless it was modified (to be constant, say) at values below 1000 mb and above 10 mb. As a practical matter, this is of no concern.

The problem of actually computing on (a portion of) the sphere has been circumvented by using an approximating plane. When the distances involved are no greater than about 0.45 radians, (according to [6], blocks of horizontal extent of no more that 18° latitude are used in NOGAPS, depending on the data density and location on the globe, with the longitudinal extent being a similar distance) there is probably no problem using great circle distance. A few calculations were performed by simply replacing the pseudoradian distance with great circle distance. In 6000 repetitions of the random process described above (for the 12Z time), no negative eigenvalues where found.

6. Summary and Conclusions

In this study innovation data over a two-month period from NOGAPS was analyzed. In addition to fitting the covariance data to estimate statistical parameters, a study was made of several weighting schemes and two procedures for fitting covariance data. The approximating function of choice was the second order autoregressive function, SOAR The possibility of including a non-negative additive constant was investigated. Based on the properties of the derived prediction and observation error covariances, and the behavior of the parameters of the approximations, it is concluded that equal weighting of the binned covariance data does not yield reliable results. Further, the use of an additive constant in the approximating function for the horizontal has a harmful effect on the positive-definiteness properties if

allowed to vary with pressure level. While this study used a weight function that cutoff at a distance of 0.2 radians (about 1275 km), the results using a cutoff distance of 0.4 radians would probably have been very similar.

The weighting scheme of choice in this study was to use a weighting for data in each bin that reflected the number of days of station pairs that contributed to that bin. In retrospect, this is not quite what was done when computing the location of the data in the bin. Here, the location was taken as the average of the abscissa and ordinate of the covariance data from each pair of stations at the distance corresponding to the interval for that bin. Subsequently, the total number of station pair days was used as the weight for fitting the SOAR function. Somewhat different data is obtained if the location of the data in the bin is computed by weighting the station pair data by the number of days for which data was collected. It is recommended that any additional studies also incorporate this weighting.

The correlation distance parameter for the horizontal approximations pose some problem. The variation with vertical position (pressure level) seems rather more pronounced and variable than seems reasonable. This could be due to the time period for the data being too short. Whether the cubic approximation used here is reasonable, and how the upper levels, for which the data is surely suspect, should be treated is not resolved satisfactorily. It is also unknown how much the correlation distances can vary over the vertical and still maintain a positive definite function.

The vertical approximation used here uses the innovative work started by Sampson and Guttorp [14], incorporating a domain transformation into the approximation. This appears to be very worthwhile. Whether or not the additive constant should be used in the vertical approximation is still an open question. There is considerable scatter in the "large-distance" correlations of both prediction and observation error, and these do not appear to settle around zero, but rather some positive number. Without the additive constant the effect would probably be to widen the correlation function to compensate for the points above the curve at the extremes (see Figure 10).

One further recommendation would be to compute using great circle distances in future studies. As long as a sufficiently limited portion of the sphere is used, the possible problems of planar versus spherical covariance functions will probably not pose any problem.

One of the more important results of this study is given by Figures 18 and 19. Note again, that the relative positions of the prediction and observation error standard deviations of the values obtained by analysis of NOGAPS data and that incorporated into the data assimilation process used in NOGAPS, are reversed at levels up to 70 mb. This means that the NOGAPS objective analysis may be treating the predicted values at levels below 70 mb as being of lesser quality than they really are, relative to the radiosonde observations. Making this change to the analysis procedure may fundamentally alter the resulting analyses. Whether or not this would lead to improved predictions is not obvious.

This study was purely univariate, using only the predicted height minus the observed heights of pressure levels used by NOGAPS. In practice, winds and temperatures need to be incorporated into simultaneous estimation of the parameters for the data assimilation process. Some experiments have been conducted in simultaneously fitting height error correlation and thickness error correlation, but no extensive study has been made.

Acknowledgments

This research was conducted while the author was a Visiting Scientist at NRL/Monterey during FY97, supported under Program Element 0602435N. The author expresses his thanks to Ed Barker, who arranged the visit, and Nancy Baker, Roger Daley, and Andrew Van Tuyl, all of whom contributed to the effort by supplying data and many useful comments.

References

1. Edward H. Barker, Design of the Navy's multivariate optimum interpolation analysis system, Weather and Forecasting 7(1992)220-231.
2. Peter Bartello and Herschel L. Mitchell, A continuous three-dimensional model of short-range forecast error covariances, Tellus 44A(1992)217-235.
3. Roger Daley, Estimating model-error covariances for application to atmospheric data assimilation, Monthly Weather Review 120(1992)1735-1746.
4. Arlindo DaSilva, Christopher Redder, and Dick Dee, Modeling retrieval error covariances for global data assimilation, preprints, Eight Conference on Satellite Meteorology, Atlanta, GA, 28 January-2 February, 1996.
5. Richard Franke, Three Dimensional Covariance Functions: Theory, Naval Research Laboratory/Monterey Memorandum Report#NRL/MR/7531-97-7231, October 1997.
6. James S. Goerss and Patricia A. Phoebus, The Navy's operational atmospheric analysis, Weather and Forecasting 7(1992)232-249.
7. A. Hollingsworth and P. Lönnberg, The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field, Tellus 38A(1986)111-136.
8. P. R. Julian and H. J. Thiébaux, On some properties of correlation functions used in optimum interpolation, Monthly Weather Review 103(1975)605-616.
9. P. Lonnberg and A. Höllingsworth, The statistical structure of short-range forecast errors as determined from radiosonde data. Part II: The covariance of height and wind errors, Tellus 38A(1986)137-161.
10. Herschel L. Mitchell, Cécilien Charette, Clément Chouinard, and Bruce Brasnett, Revised interpolation statistics for the Canadian data assimilation procedure: their derivation and application, MWR 118(1990)1591-1614.
11. Herschel L. Mitchell, Clément Chouinard, and Cécilien Charette, Impact of a revised analysis algorithm on operational data assimilation system, Monthly Weather Review 124(1996)1243-1255.
12. M. B. Priestley, Spectral Analysis and Time Series, Vol. 1: Univariate Series, and Vol. 2: Multivariate Series, Prediction and Control, Academic Press, 1981.
13. Florence Rabier and Tony McNally, Evaluation of forecast error covariance matrix, Technical Memorandum No. 195, ECMWF, 1993.
14. Paul D. Sampson and Peter Guttorp, Nonparametric estimation of nonstationary spatial covariance structure, J. American Statistical Society 87(1992)108-119.

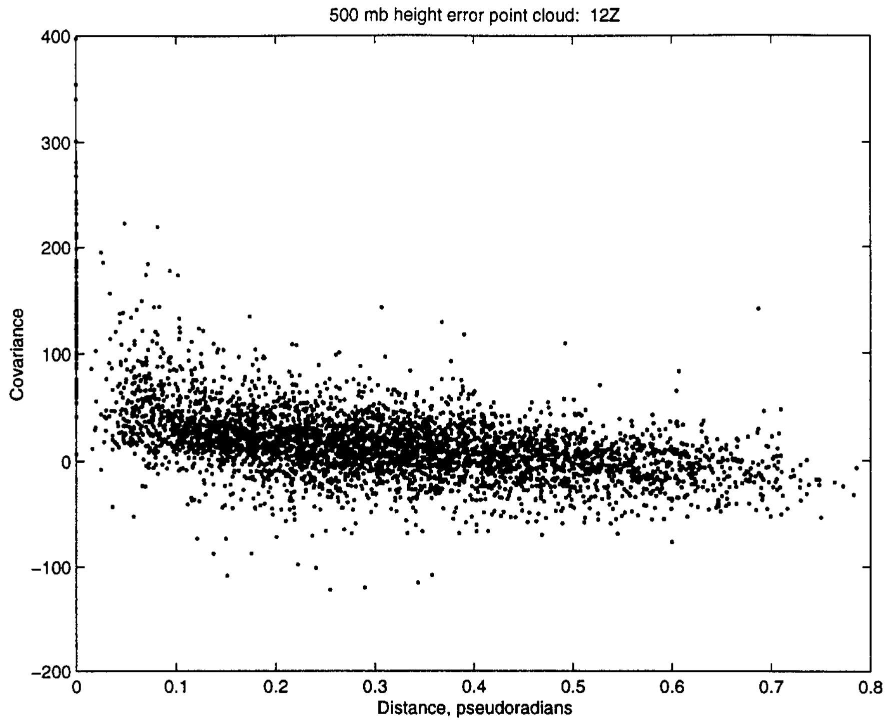

Figure 1: Inter-station covariances, 500 mb at 12Z.

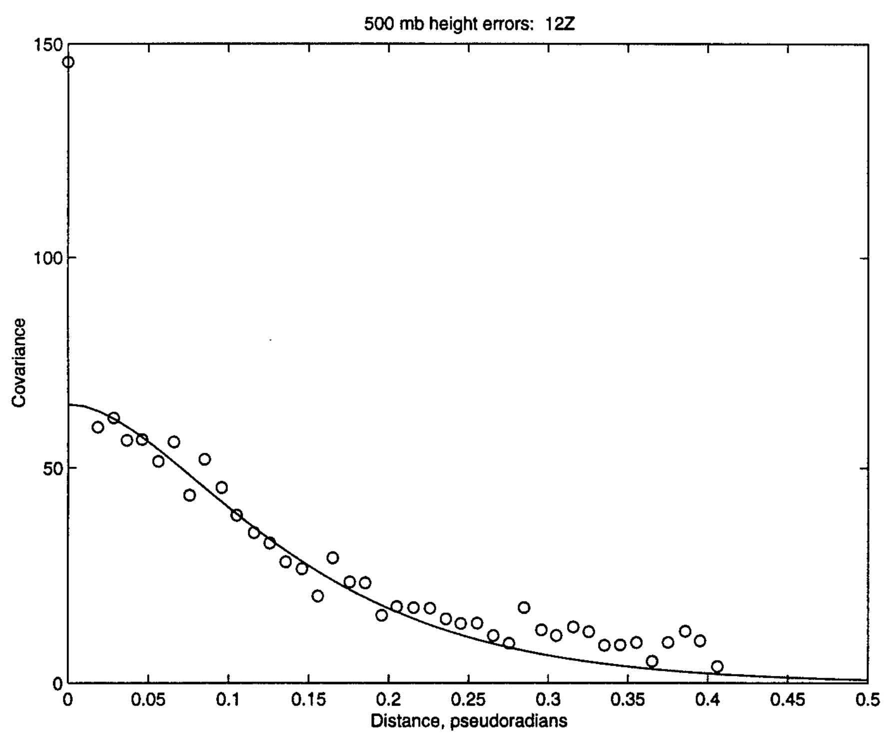

Figure 2: Binned covariances and curve for 500 mb height errors at 12Z.

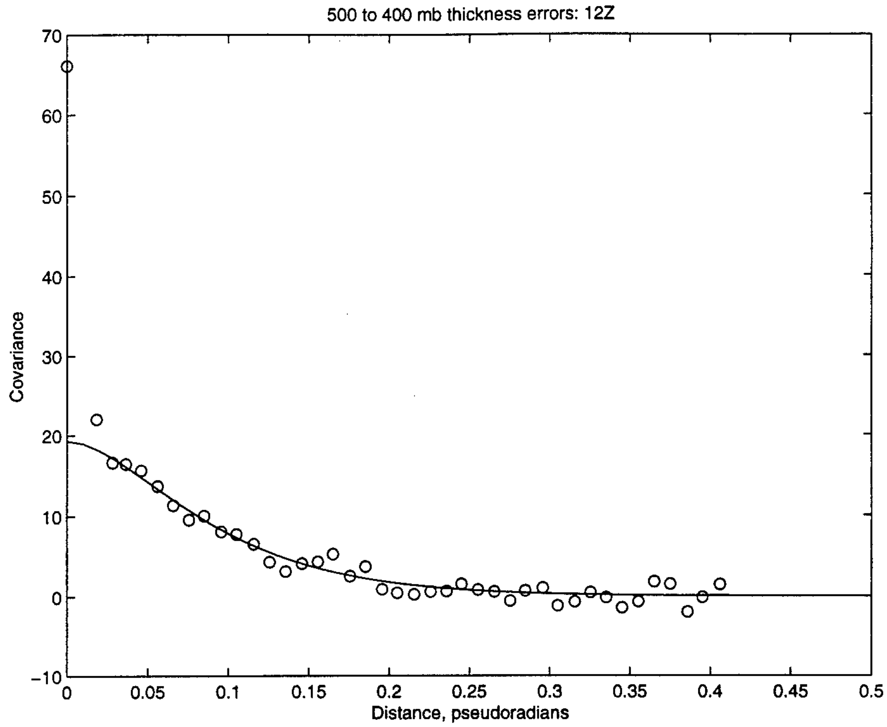

Figure 3: Binned covariances and curve for 500 to 400 mb thickness errors at 12Z.

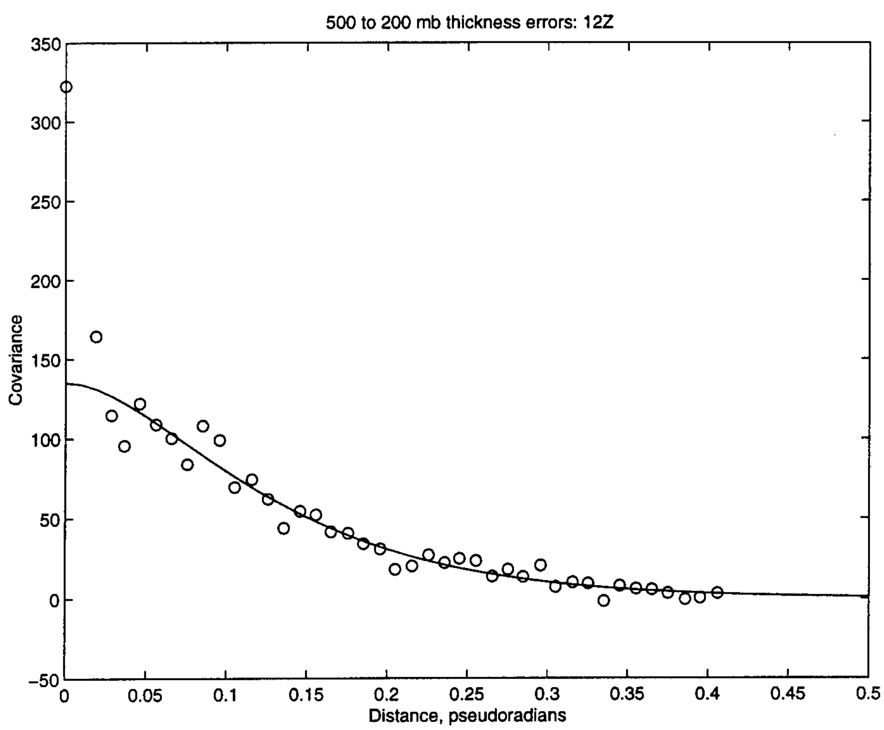

Figure 4: Binned covariances and curve for 500 to 200 mb thickness errors at 12Z.

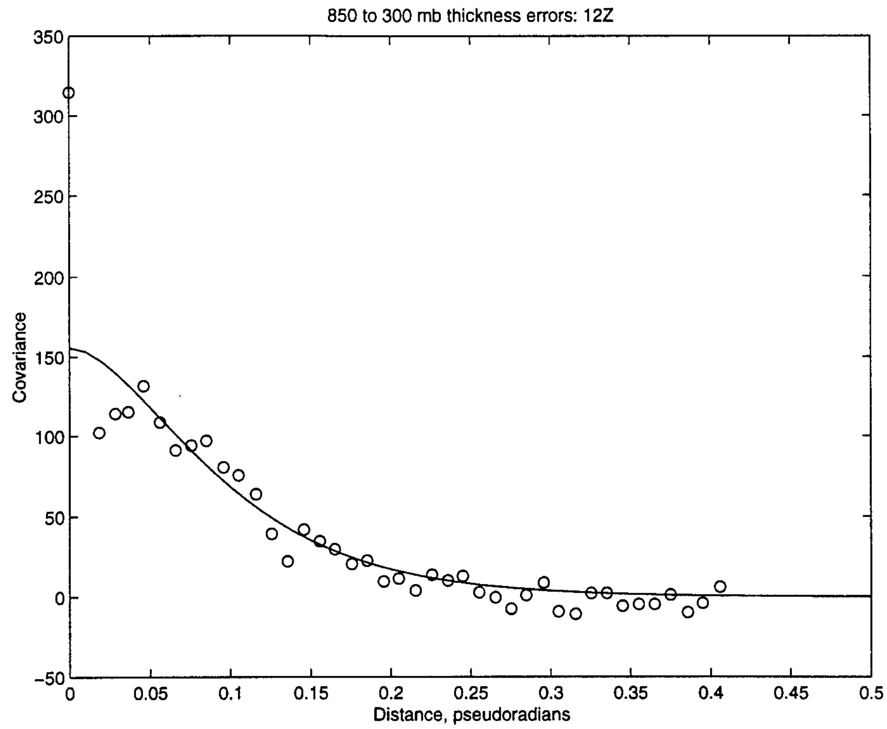

Figure 5: Binned covariances and curve for 850 to 300 mb thickness errors at 12Z.

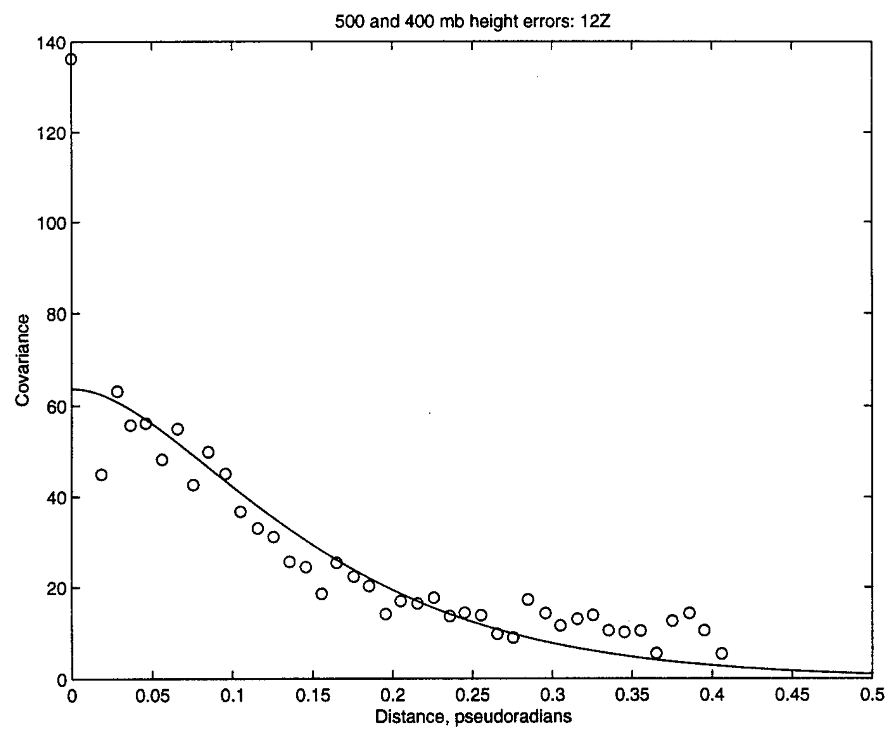

Figure 6: Binned covariances and implied curve for 500 and 400 mb height errors at 12Z.

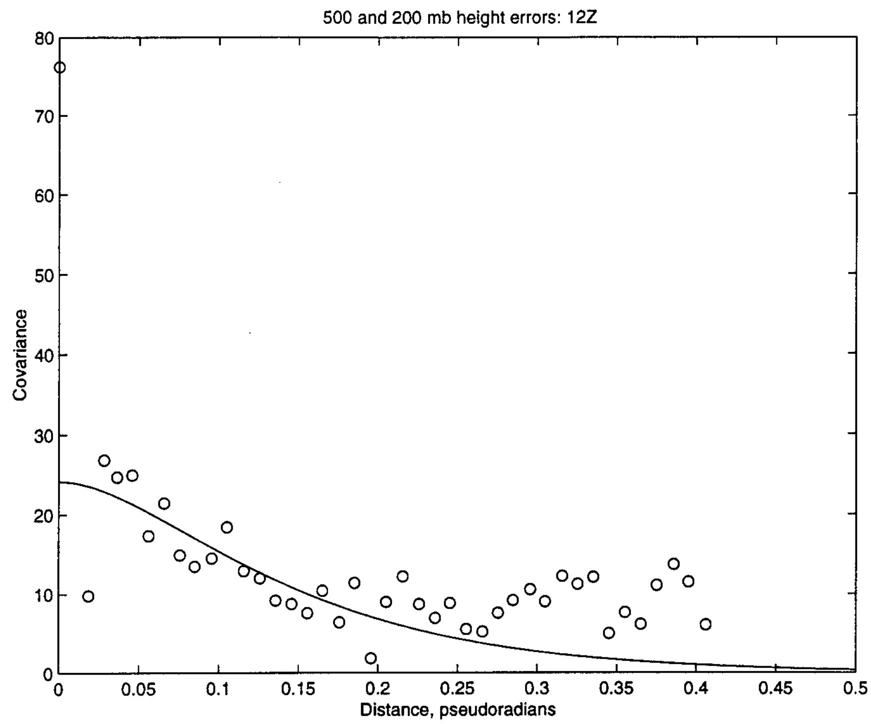

Figure 7: Binned covariances and implied curve for 500 and 200 mb height errors at 12Z.

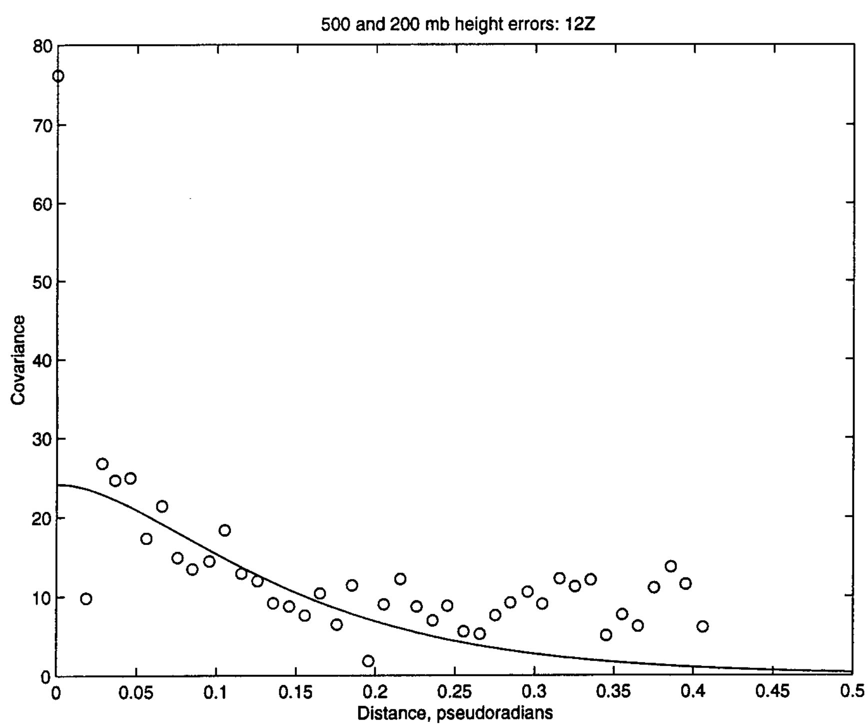

Figure 8: Binned covariances and implied curve for 850 and 300 mb height errors at 12Z.

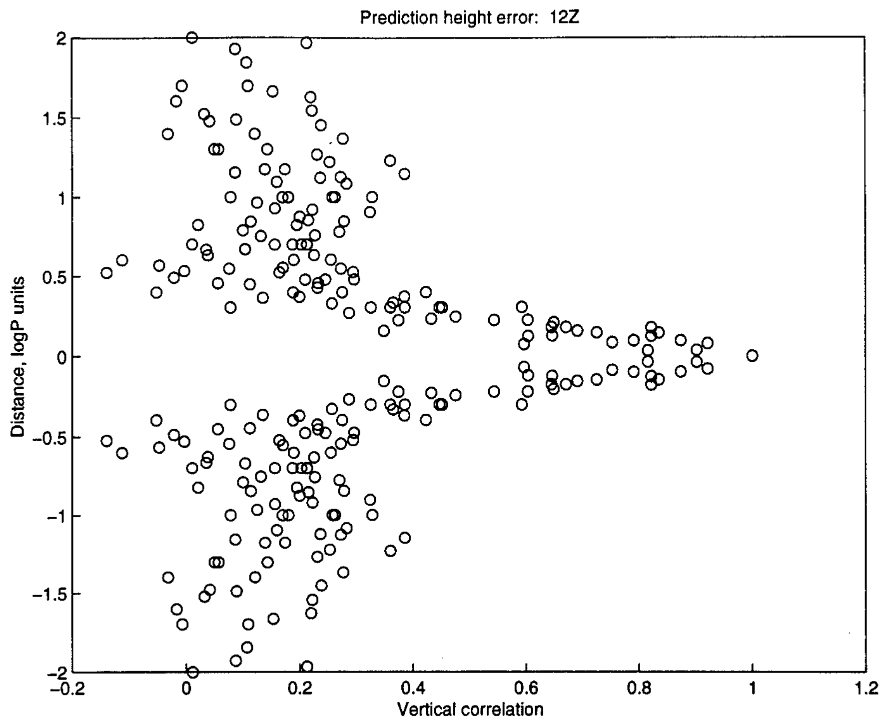

Figure 9: Vertical correlation of prediction error in logP distance at 12Z.

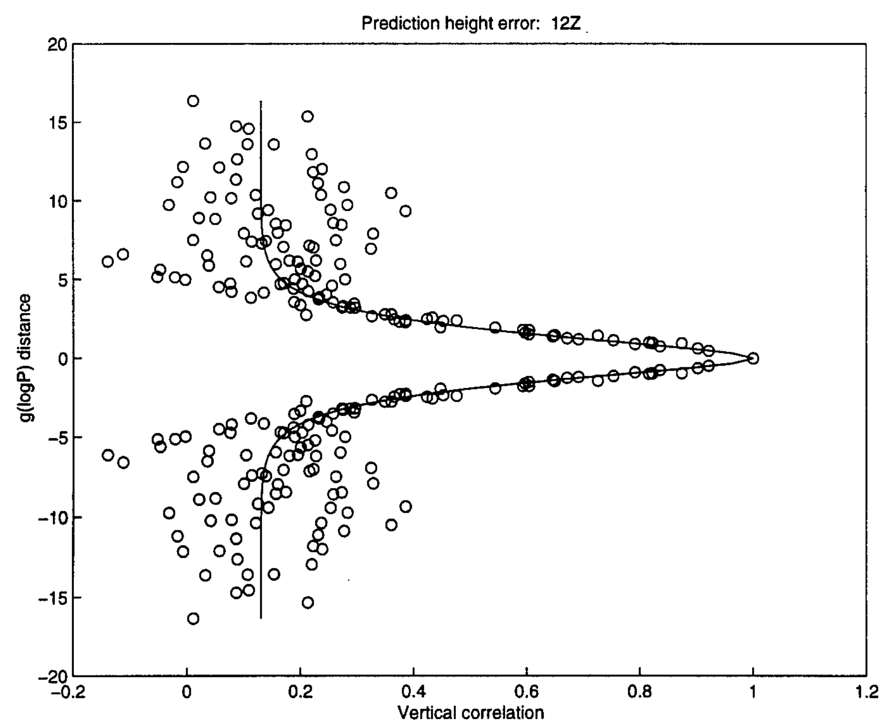

Figure 10: Vertical correlation of prediction error in transformed coordinate distance, with curve, at 12Z.

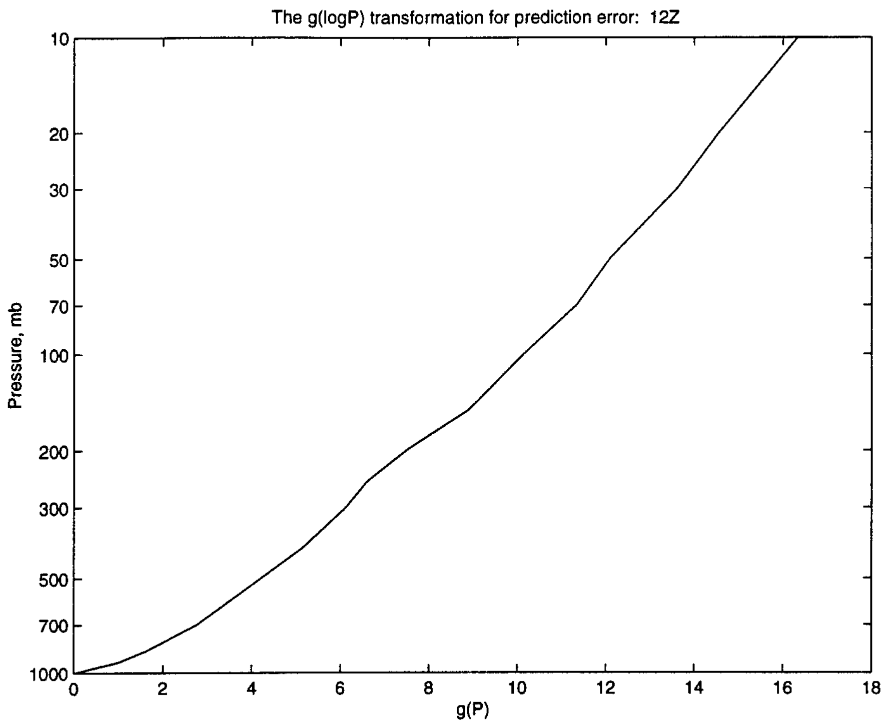

Figure 11: Distance transformation for prediction error correlation in the vertical coordinate at 12Z.

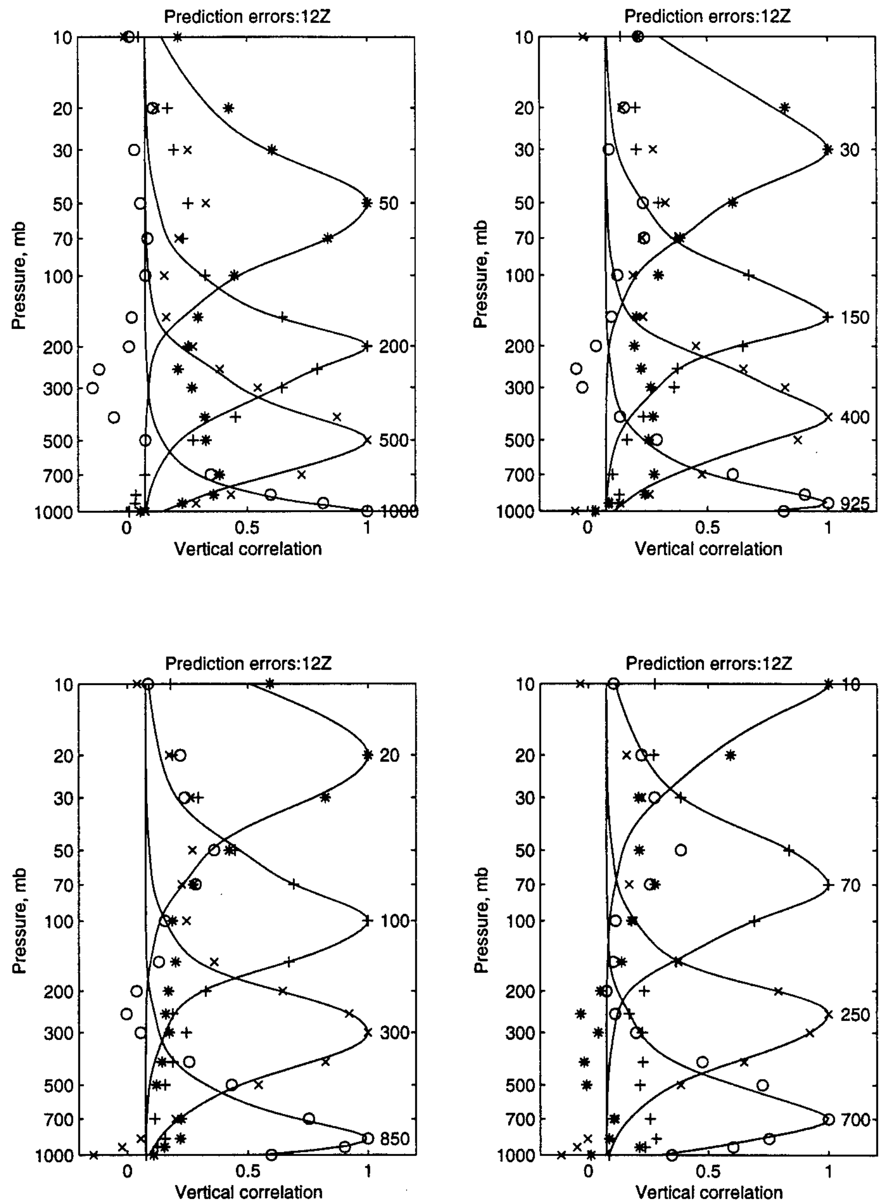

Figure 12: Vertical correlation of prediction error, with curves, at 12Z.

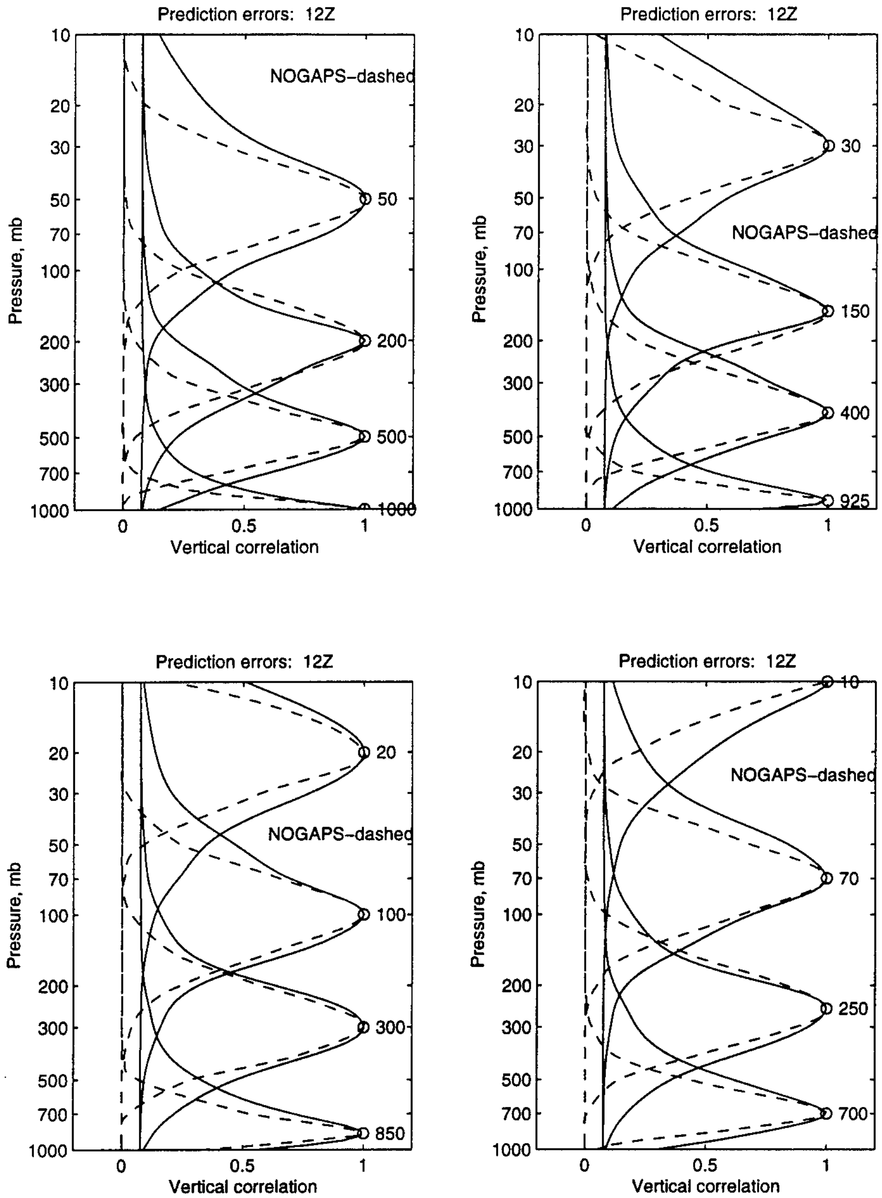

Figure 13: Vertical correlation curves for prediction error, with NOGAPS curves, at 12Z.

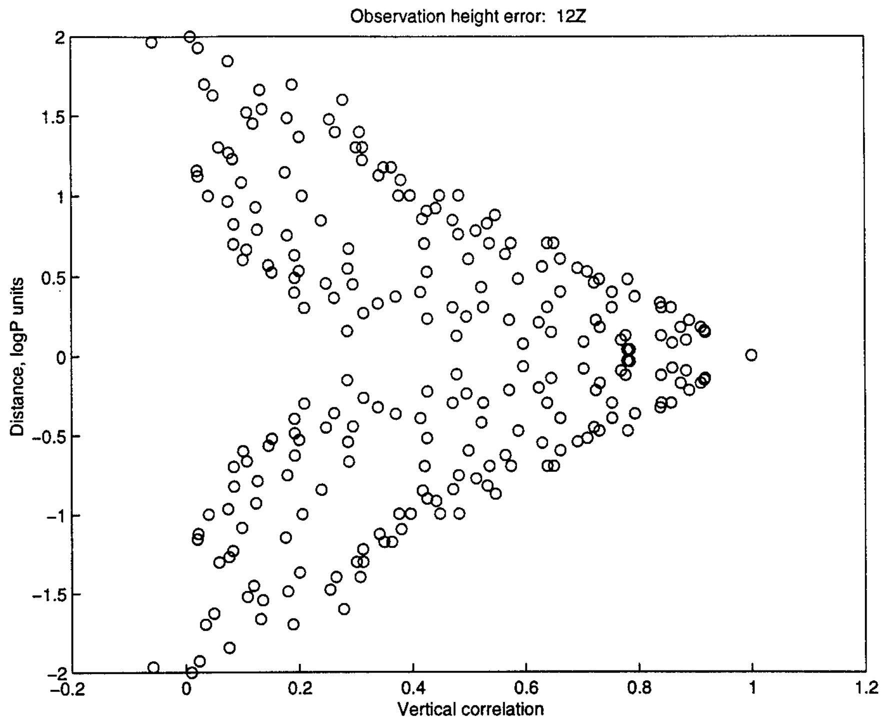

Figure 14: Vertical correlation of observation error in logP distance at 12Z.

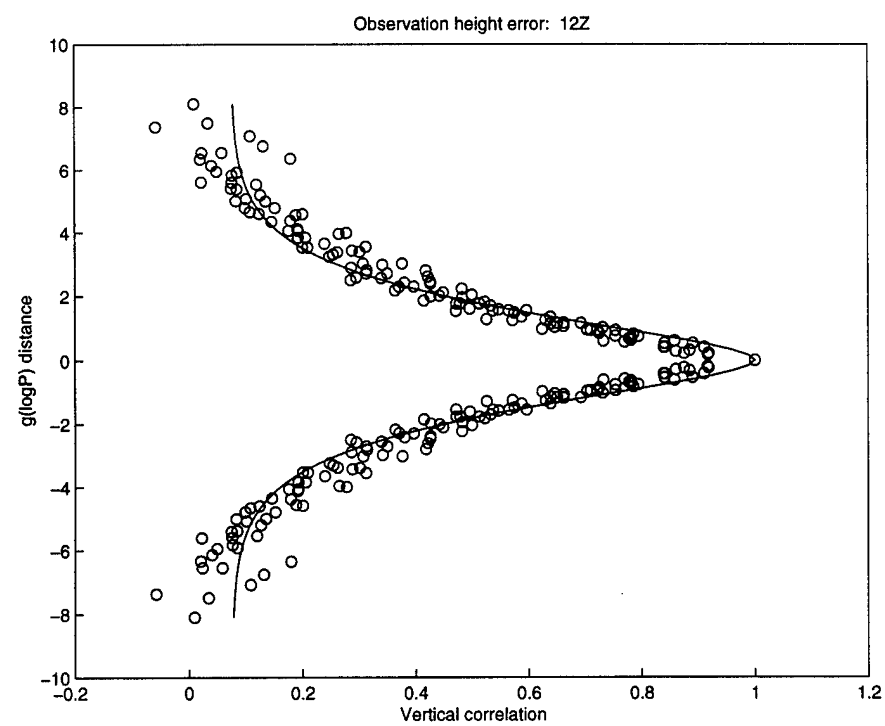

Figure 15: Vertical correlation of observation error in transformed coordinate distance, with curve, at 12Z.

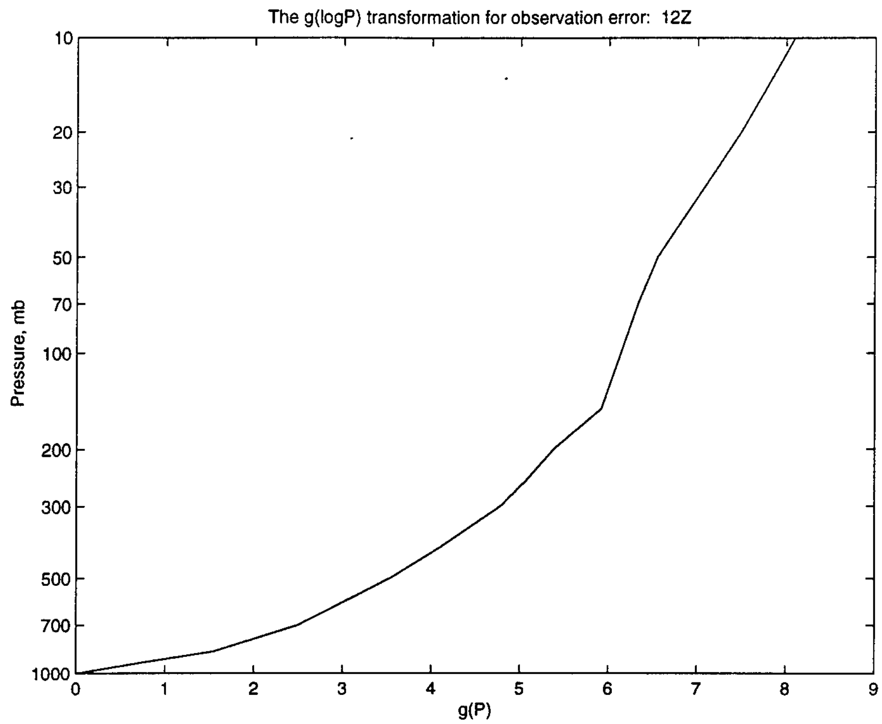

Figure 16: Distance transformation for observation error correlation in the vertical coordinate at 12Z.

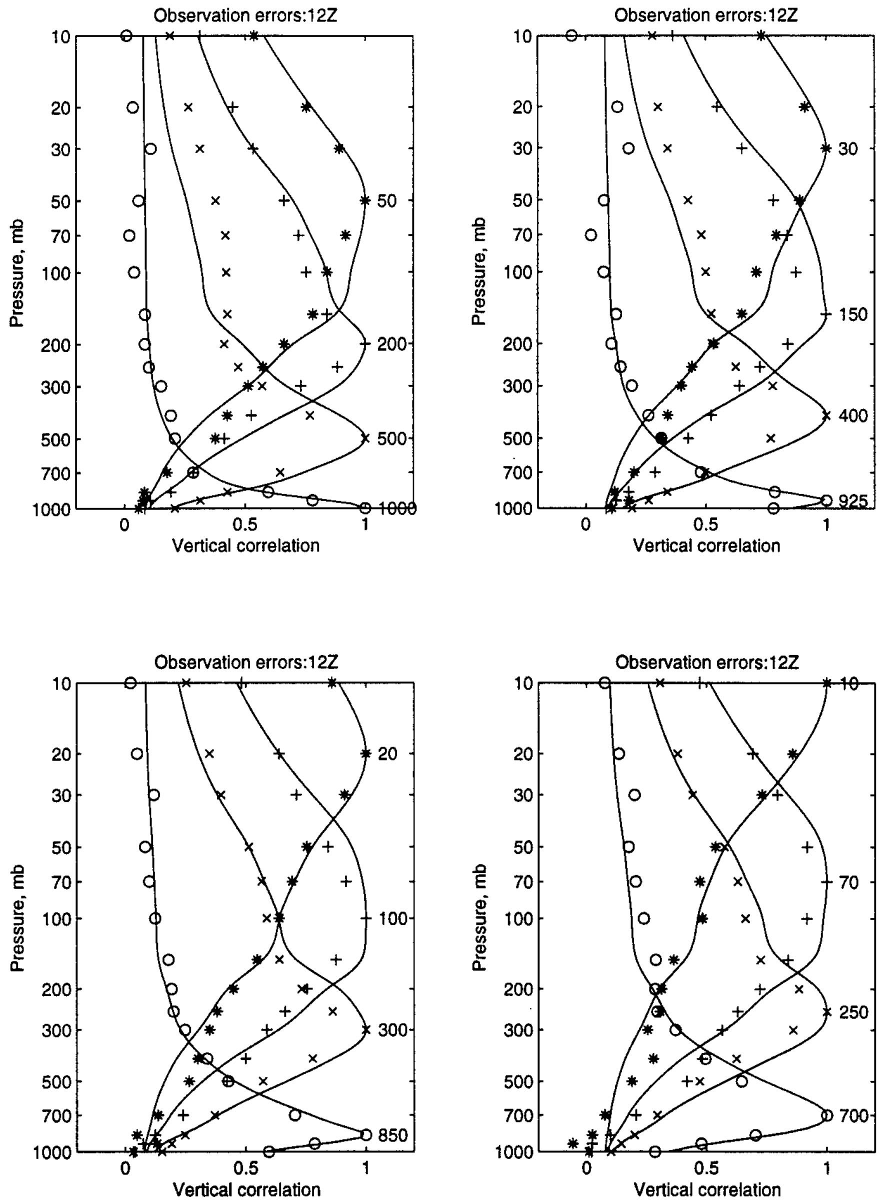

Figure 17: Vertical correlation of observation error, with curves, at 12Z.

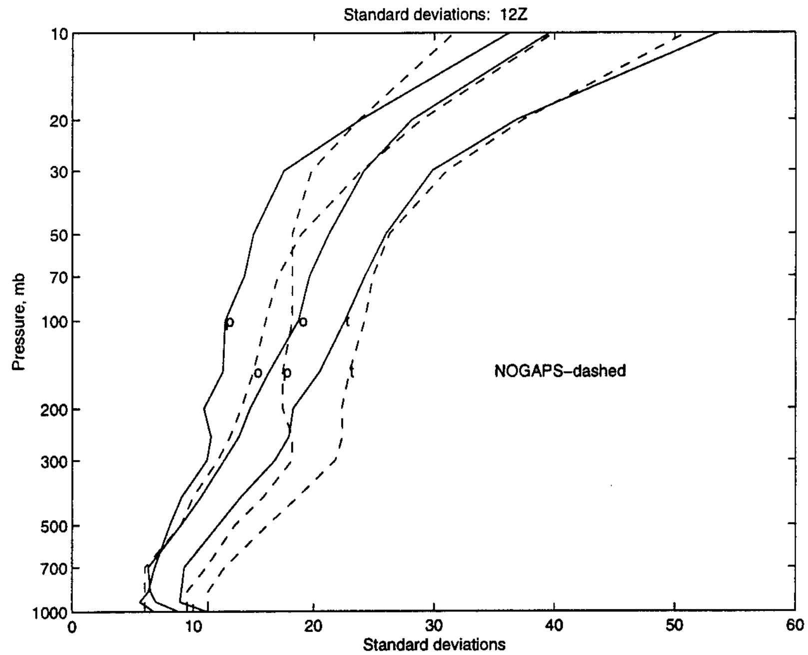

Figure 18: Standard deviations of prediction, observation, and total errors at 12Z, with the values presently assumed in NOGAPS.

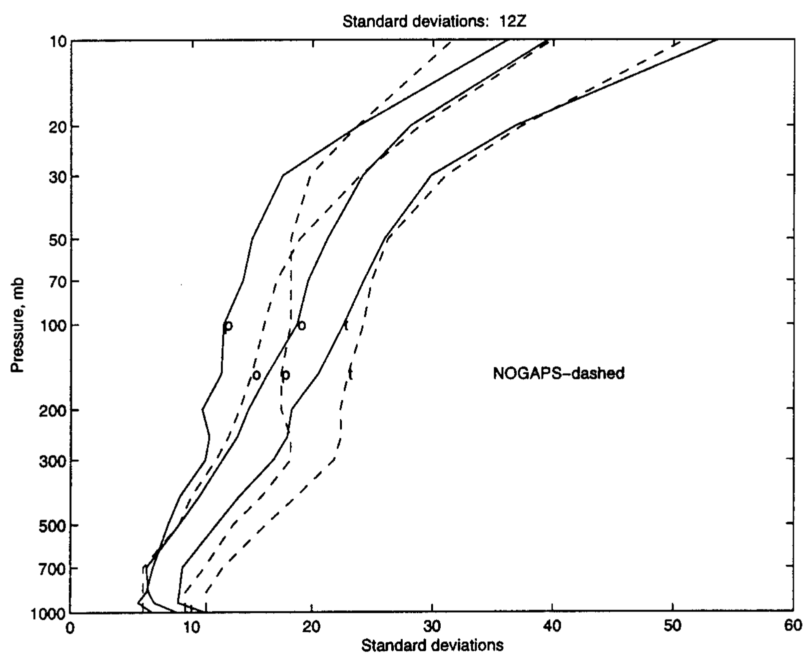

Figure 19: Standard deviations of prediction, observation, and total errors at 00Z, with the values presently assumed in NOGAPS.

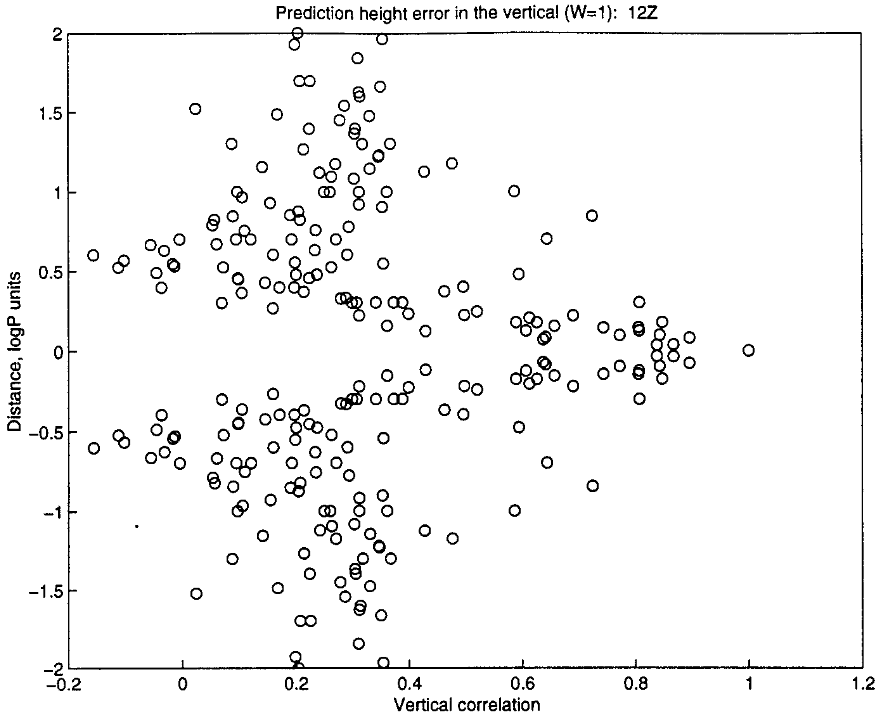

Figure 20: Vertical correlation of prediction error in logP distance at 12Z, using $W_n = 1$.

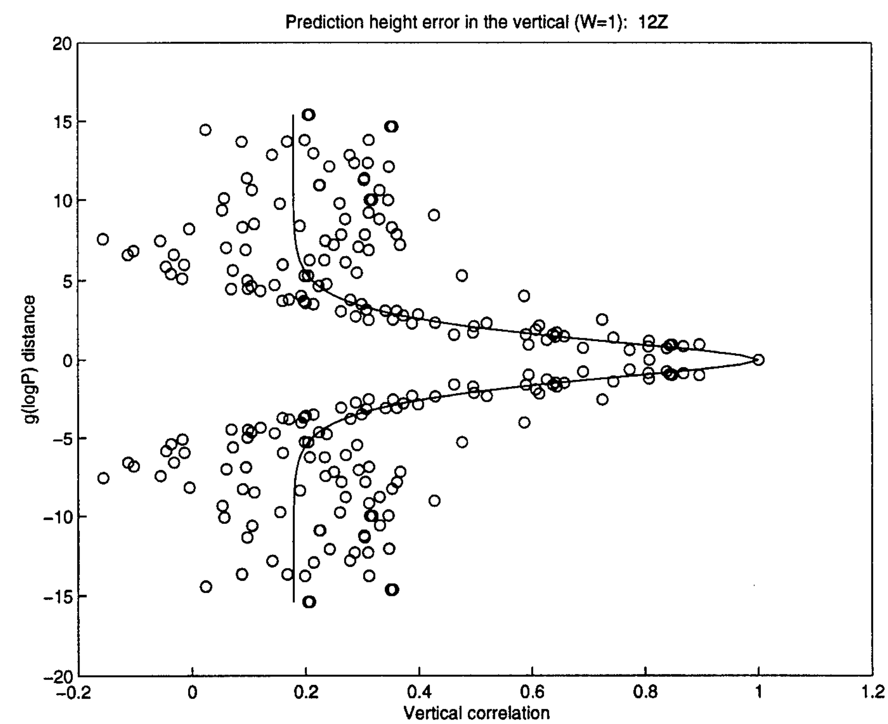

Figure 21: Vertical correlation of prediction error in transformed coordinate distance, with curve, at 12Z, using $W_n = 1$.

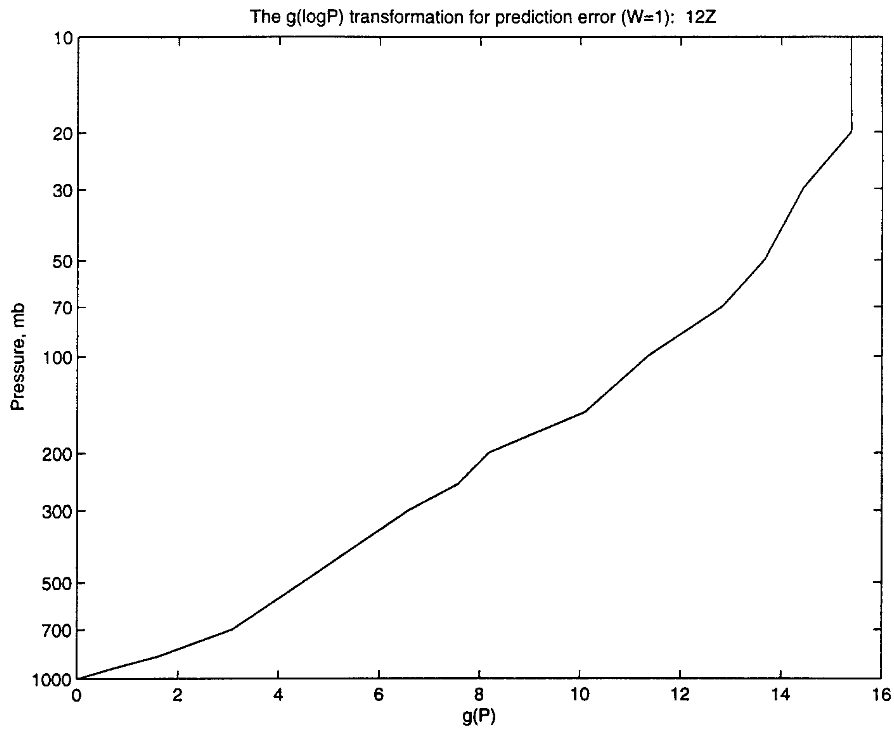

Figure 22: Distance transformation for prediction error correlation in the vertical coordinate at 12Z using $W_n = 1$.

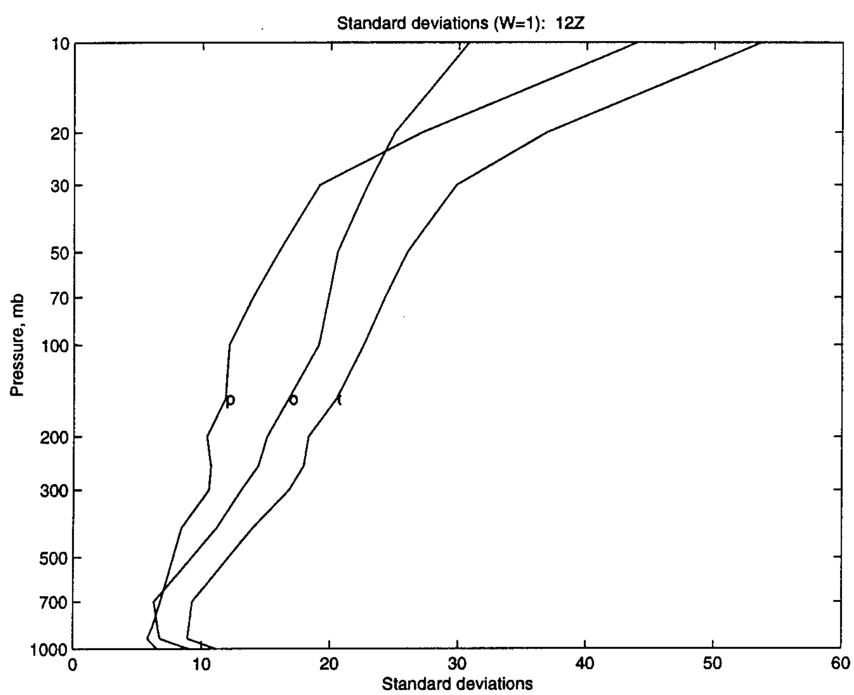

Figure 23: Standard deviations of prediction, observation, and total error aat 12Z using $W_n = 1$.

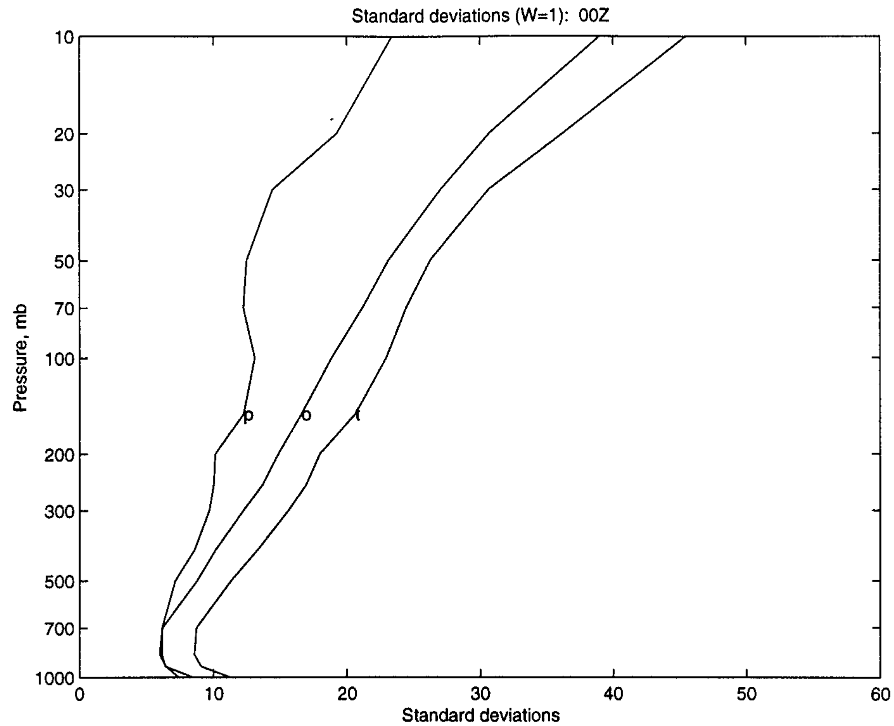

Figure 24: Standard deviations of prediction, observation, and total error aat 00Z using $W_n = 1$.

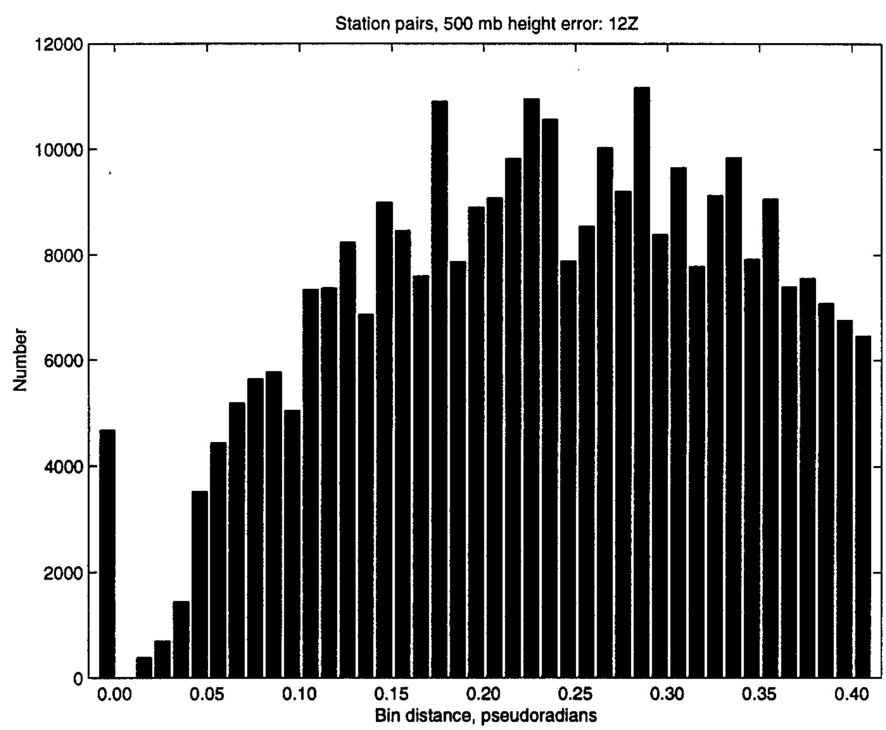

Figure 25: Number of products of prediction error in each bin for the 500 mb level, at 12Z.

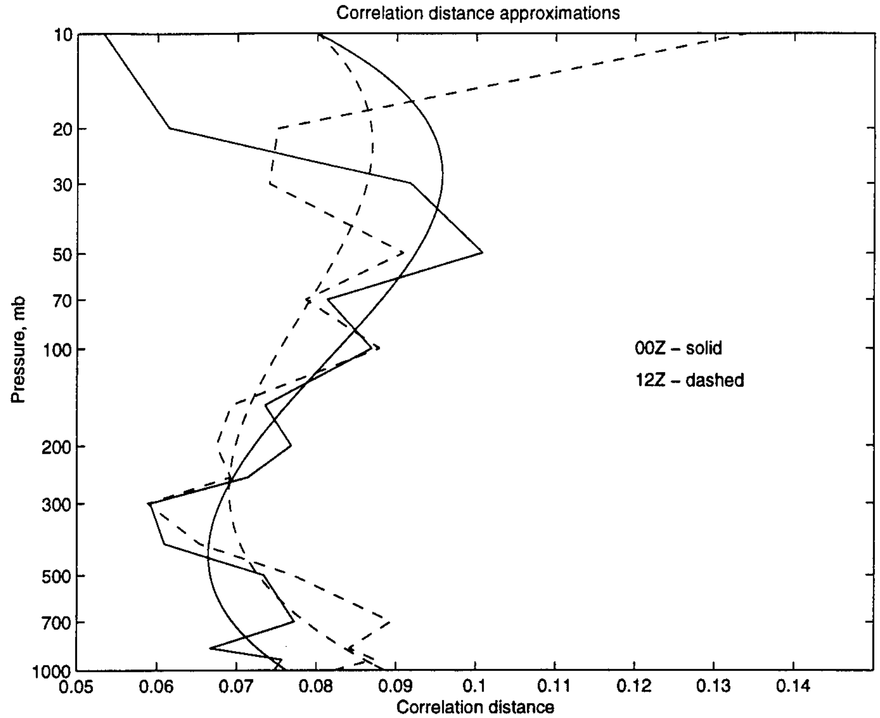

Figure 26: Correlation distance for the levels, and cubic approximations.

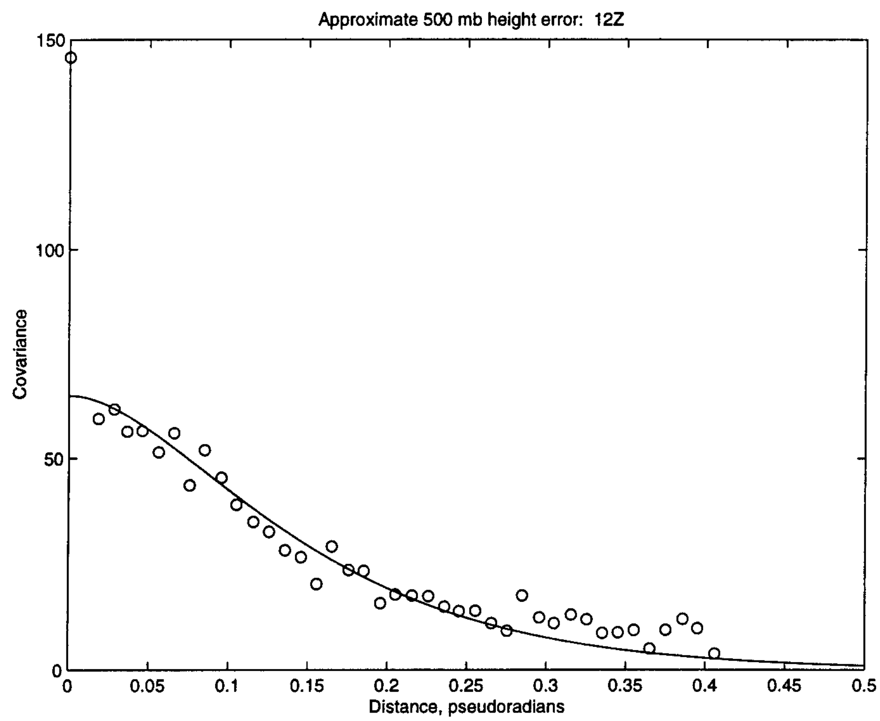

Figure 27: Binned covariances and approximate curve for 500 mb height errors at 12Z.

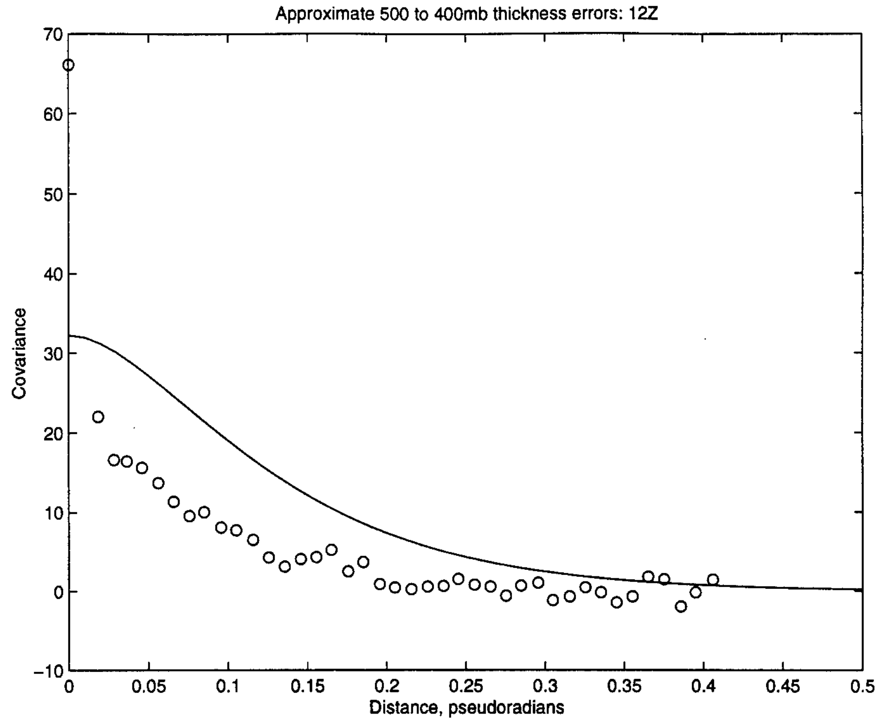

Figure 28: Binned covariances and approximate curve for 500 to 400 mb thickness errors at 12Z.

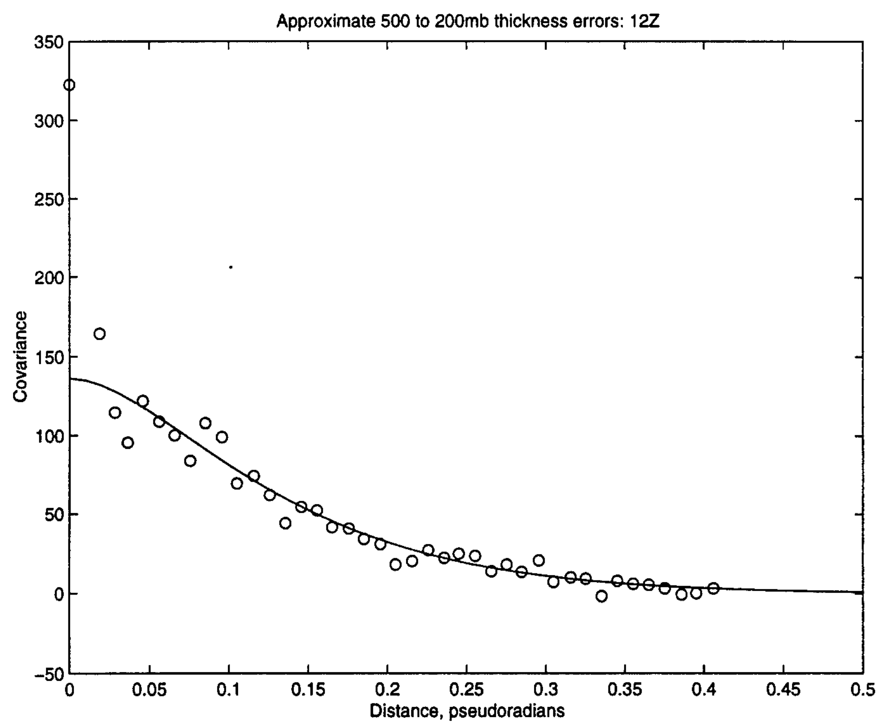

Figure 29: Binned covariances and approximate curve for 500 to 200 mb thickness errors at 12Z.

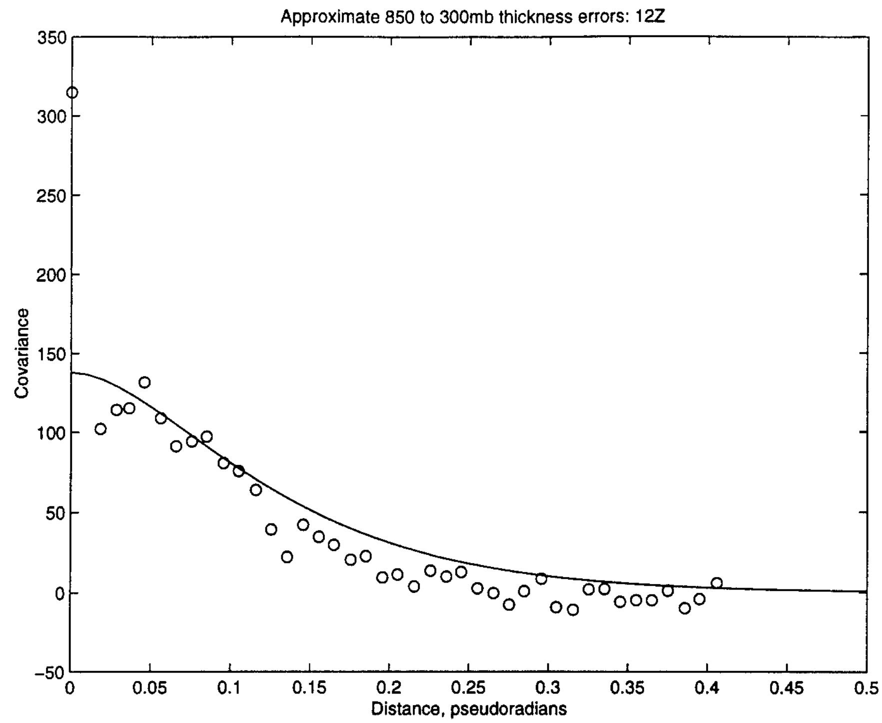

Figure 30: Binned covariances and approximate curve for 850 to 300 mb thickness errors at 12Z.

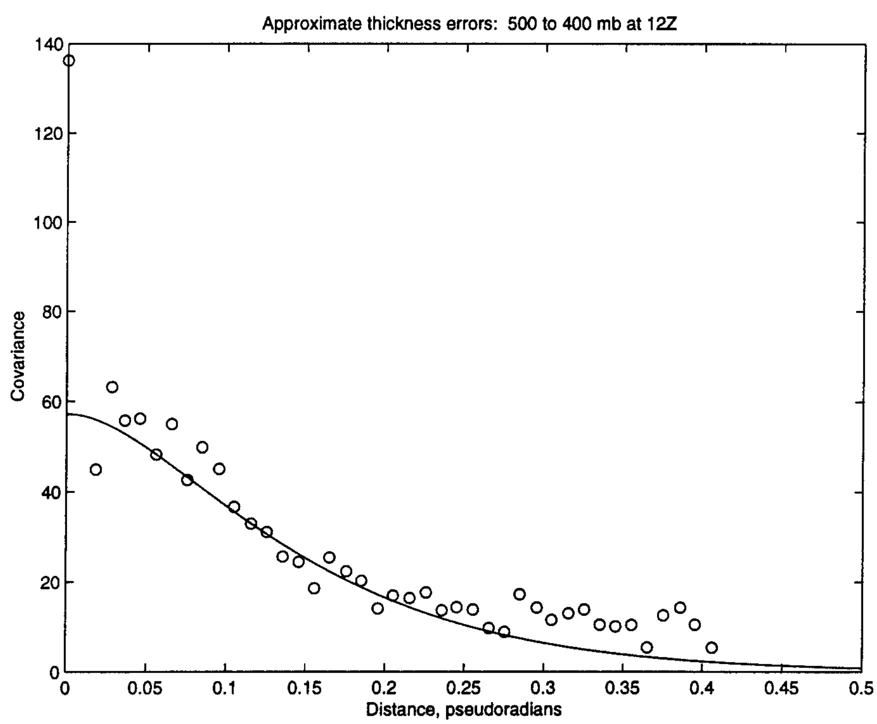

Figure 31: Binned covariances and approximate curve for 500 and 400 mb height errors at 12Z.

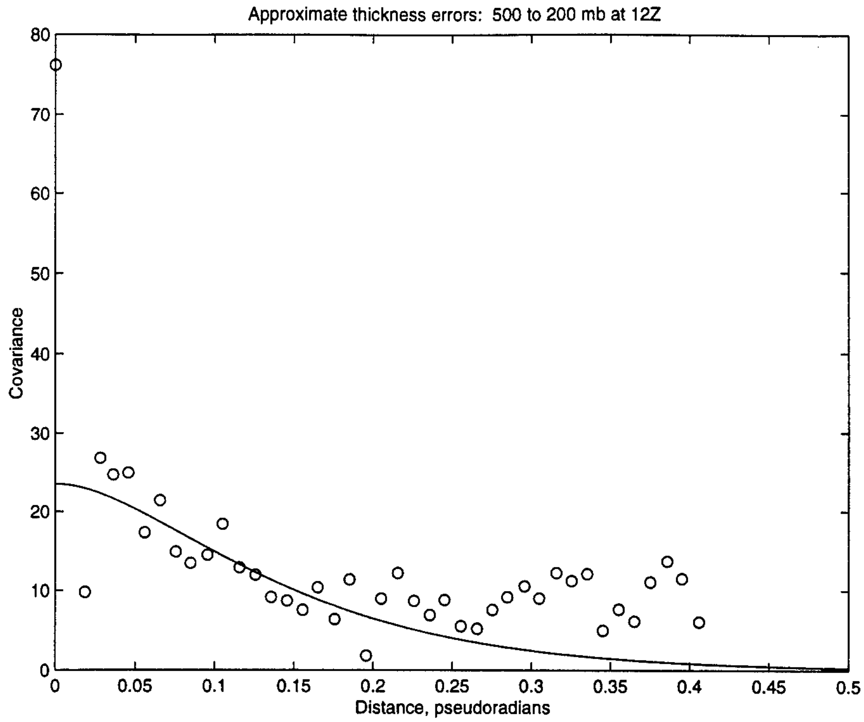

Figure 32: Binned covariances and approximate curve for 500 and 200 mb height errors at 12Z.

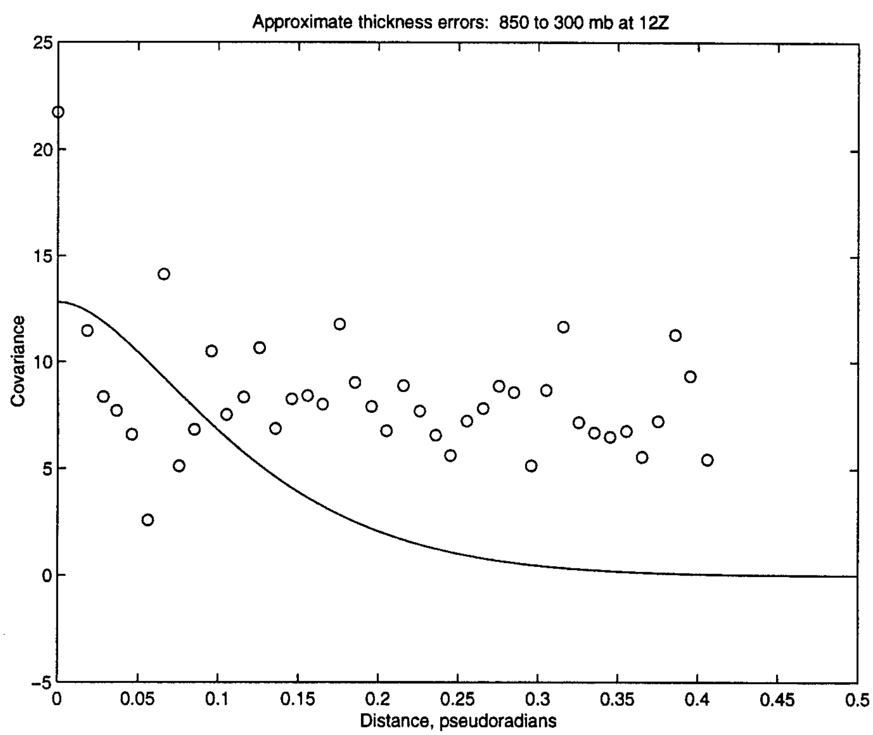

Figure 33: Binned covariances and approximate curve for 850 and 300 mb height errors at 12Z.

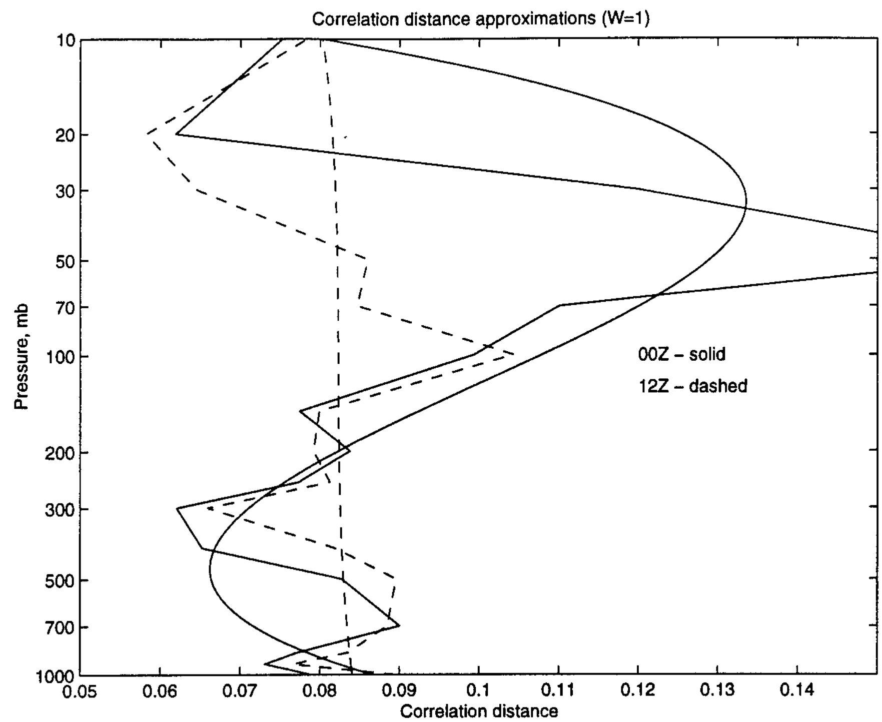

Figure 34: Correlation distance for the levels, and cubic approximations using W_n=1.

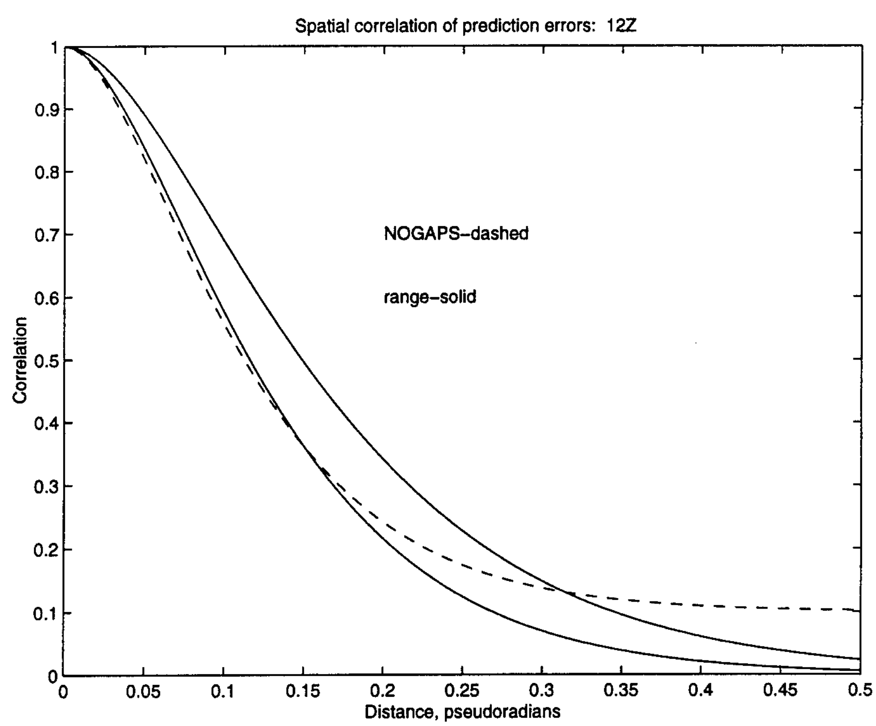

Figure 35: Spatial correlation functions with extreme correlation distances, and NOGAPS.